Identification of novel genetic and prognostic markers in hereditary and sporadic cancer: “Two sides of the same coin”

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel
von

Salvatore Piscuoglio
aus Neapel, Italien

Basel 2012
Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Primo Schär
Prof. Dr. Gernot Jundt
Prof. Dr. Karl Heinimann
Prof. Dr. Luigi M. Terracciano
Prof. Alfredo Fusco

Basel, den 13-11-2012

Prof. Dr. Jörg Schibler
Dekan
“I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale.”

Marie Curie (1867 - 1934)
# 1. TABLE OF CONTENT

1. TABLE OF CONTENT ........................................................................................................ iv

2. ABBREVIATIONS ............................................................................................................. vi

3. ABSTRACT ....................................................................................................................... 1

4. GENERAL INTRODUCTION .......................................................................................... 4
   4.1 Cancer .......................................................................................................................... 4
   4.2 Cancer Genetics .......................................................................................................... 7
       4.2.1 Colorectal Cancer ............................................................................................... 10
   4.3 Hereditary cancer syndromes .................................................................................... 13
       4.3.1 Lynch syndrome (aka HNPCC) .......................................................................... 15
           4.3.1.1 Mismatch repair system .............................................................................. 19
           4.3.1.2 Microsatellite instability ............................................................................ 21
   4.4 Other hereditary colorectal cancer syndromes ......................................................... 23
       4.4.1 Familial adenomatous polyposis ....................................................................... 23
           4.4.1.1 Familial adenomatous polyposis variants .................................................. 24
       4.4.2 Hamartomatous polyposis syndromes ............................................................... 25
           4.4.2.1 Juvenile polyposis ...................................................................................... 25
           4.4.2.2 Peutz–Jeghers syndrome .......................................................................... 26
           4.4.2.3 PTEN Hamartoma Tumor Syndrome ......................................................... 27
   4.5 Diagnostic and prognostic genetic markers in cancer .............................................. 29
   4.6 MicroRNA: biogenesis, functions and targets ........................................................... 34
       4.6.1 MicroRNA and cancer ....................................................................................... 36
   4.7 HMG proteins ............................................................................................................ 38
       4.7.1 HMG A family .................................................................................................... 38
       4.7.2 HMG A expression in normal and neoplastic tissues ...................................... 40
   4.8 Cancer Stem Cell Markers ....................................................................................... 42
   4.9 MAGE Gene Family ................................................................................................ 43

5. AIMS ............................................................................................................................... 45

6. RESULTS ......................................................................................................................... 47
   6.1 3’UTR poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair deficient cancer ................................................... 47
6.2 The 8p21.3 encoded SHOCA-2 acts as a tumor suppressor in colorectal cancer via repression of STAT3 activation. ...................................................... 78
6.3 HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma. ............ 146
6.3.1 HMGA1 over-expression represents a poor prognostic index in human breast carcinoma. ................................................................. 155
6.4 Effect of EpCAM, CD44, CD133 and CD166 expression on patient survival in tumours of the ampulla of Vater .................................................. 176
6.5 MAGE-A10 is a nuclear protein frequently expressed in high percentages of tumor cells in lung, skin and urothelial malignancies. ...................... 183
7. GENERAL DISCUSSION ........................................................................................................... 196
  7.1 3’UTR poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair deficient cancers ........................................... 196
  7.2 SH2D4A as a novel tumor suppressor gene in CRCs ...................................................... 199
  7.3 HMGA proteins as prognostic markers in different tumor entities ......................... 201
  7.4 MAGE-A10 overexpression in lung, skin and urothelial malignancies .................. 203
8. APPENDIX .......................................................................................................................... 205
  8.1 microRNA expression profiling in mismatch repair associated colorectal cancer ................................................................. 205
  8.2 Identification of novel recurrent duplication “hot spots” in Lynch syndrome colorectal cancers. ............................................................... 218
9. REFERENCES ...................................................................................................................... 233
10. CURRICULUM VITAE ....................................................................................................... 254
11. ACKNOWLEDGMENTS .................................................................................................... 261
## 2. ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Amsterdam Criteria</td>
</tr>
<tr>
<td>AFAP</td>
<td>Attenuated Familial Adenomatous Polyposis</td>
</tr>
<tr>
<td>APC</td>
<td>Adenomatous Polyposis Coli</td>
</tr>
<tr>
<td>BRRS</td>
<td>Bannayan-Riley-Ruvalcaba Syndrome</td>
</tr>
<tr>
<td>BG</td>
<td>Bethesda Guidelines</td>
</tr>
<tr>
<td>CHRPE</td>
<td>Congenital hypertrophy of the retinal pigment epithelium</td>
</tr>
<tr>
<td>CS</td>
<td>Cowden Syndrome</td>
</tr>
<tr>
<td>CRC</td>
<td>Colorectal Cancer</td>
</tr>
<tr>
<td>FAP</td>
<td>Familial Adenomatous Polyposis</td>
</tr>
<tr>
<td>HE</td>
<td>Hemotoylin and Eosin Stain</td>
</tr>
<tr>
<td>hMLH</td>
<td>Human MutL Homolog</td>
</tr>
<tr>
<td>hMSH</td>
<td>Human MutS Homolog</td>
</tr>
<tr>
<td>HMPS</td>
<td>Hereditary Mixed Polyposis Syndrome</td>
</tr>
<tr>
<td>hPMS</td>
<td>Human Post Meiotic Segregation</td>
</tr>
<tr>
<td>HNPCC</td>
<td>Hereditary Non Polyposis Colorectal Cancer</td>
</tr>
<tr>
<td>IARC</td>
<td>International Agency for Research on Cancer</td>
</tr>
<tr>
<td>IDLs</td>
<td>Insertion or deletion loops</td>
</tr>
<tr>
<td>IHC</td>
<td>Immunohistochemistry</td>
</tr>
<tr>
<td>KRAS</td>
<td>Kirsten rat sarcoma viral oncogene homolog</td>
</tr>
<tr>
<td>LOH</td>
<td>Loss of Heterozygosity</td>
</tr>
<tr>
<td>MMR</td>
<td>Mismatch Repair</td>
</tr>
<tr>
<td>MSI</td>
<td>Microsatellite Instability</td>
</tr>
<tr>
<td>MYH</td>
<td>mutY homolog</td>
</tr>
<tr>
<td>PJS</td>
<td>Peutz-Jeghers Syndrome</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PTEN</td>
<td>Phosphatase and Tensin Homologue</td>
</tr>
<tr>
<td>Rb</td>
<td>Retinoblastoma gene</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>TGFβRII</td>
<td>Tumour Growth Factor βeta Receptor II</td>
</tr>
<tr>
<td>TP53</td>
<td>Tumor protein 53</td>
</tr>
<tr>
<td>JPS</td>
<td>Juvenile Polyposis Syndrome</td>
</tr>
</tbody>
</table>
3. ABSTRACT

This thesis has focused on the discovery and characterization of novel diagnostic and prognostic markers in various cancer entities, with a special emphasis on colorectal cancer (CRC).

In Switzerland the incidence of colorectal cancer ranks third in males and second in females, with about 4000 new patients diagnosed each year. Incidence trends over the last decades have remained constant in both sexes, whereas mortality rates have been decreasing. Decreasing mortality is thought to be related to improved treatment during the past years as well as generalisation of colorectal cancer screening in the Swiss population. About 80% of colorectal cancers are thought to have occurred by chance (sporadic) with the remainder displaying either familial aggregation (about 15%) or mendelian inheritance (about 5%).

In the first part of this work we identify and characterize a novel target gene locus for microsatellite instability (MSI) consisting of a mononucleotide (T/U)16 tract, EWS16T, located in the 3' UTR of the Ewing sarcoma break point region 1 (EWSR1) gene in 319 patients with hereditary and sporadic CRC. We show that the EWS16T locus discriminates MMR proficient from deficient cancers with high diagnostic sensitivity (100%) and specificity (100%). It could thus substantially improve and facilitate MSI analysis in routine daily practice. In addition, biochemical analyses indicate that EWS16T contractions alter poly(A) site selection by promoting SFPQ-mediated distal poly(A) site usage in EWSR1 pre-mRNAs and result in decreased mRNA as well as EWS protein expression. Our findings thus directly implicate the RNA-/DNA-binding Ewing sarcoma protein in MSI-associated colorectal tumorigenesis.

In the second part we characterize a new tumour suppressor gene designated SH2D4A located on the short arm of chromosome 8. We demonstrate that SH2D4A physically interacts with the EGFR/STAT3 pathway and controls cell proliferation. Upon EGF signaling, SH2D4A recruits the serine/threonine phosphatase PP1β to the receptor complex and represses activated STAT3 via dephosphorylation. SH2D4A expression reduces anchorage-independent tumour cell growth and its loss promotes the expression of c-Myc, Cyclin D1
and Jun B. In addition we show that $SH2D4A$ expression is partially lost in human colorectal cancers as a result of chromosomal instability, mutations and epigenetic changes. Finally, diminished $SH2D4A$ protein expression was found to correlate with advanced disease stages and was associated with poor prognosis.

In the third part we investigate HGMA1/HGM2 protein expression 210 and 1202 patients with pancreatic and breast cancers, respectively. HMGA1 and HMGA2 over-expression was found in a significant number of breast and pancreatic carcinoma samples, and its over-expression positively correlated with grade and stage of the disease. Conversely, no HMGA1 and HMGA2 expression was observed in cancer-free breast and pancreas tissues. Taken together, our findings show that high expression levels of HMGA1 and HMGA2 are related to an unfavorable histological type and a poor prognosis in both, pancreatic and breast cancer. Moreover, these findings further support the notion that these proteins represent appropriate targets for the therapy of human cancer, as suggested by numerous in vitro and in vivo studies.

In the fourth part of the thesis we evaluate the potential role of the cancer stem cell (CSC) proteins EpCAM, CD44s, CD166 and CD133 in tumors of the ampulla of Vater. CSC expression was determined in 175 carcinoma, 111 adenoma and 152 cancer free-mucosa specimens arranged on a tissue microarray format. The expression of all evaluated marker proteins differed significantly between cancer-free mucosa, adenoma and carcinoma samples. EpCAM expression was significantly correlated with better patient survival. In contrast, increased expression of CD44s, CD166 and CD133 from normal mucosa samples to adenoma and carcinoma was linked to tumor progression but no statistically significant correlation with survival observed. Our findings therefore indicate that in ampullary carcinomas loss of EpCAM expression may be associated with a more aggressive tumor phenotype.

In the fifth part we develop a specific monoclonal antibody for the highly immunogenic member of the MAGE-A family of cancer/testis tumor-associated antigens (C/T TAAs). The antibody was used to stain a multi-tumor tissue microarray comprising more than 2,500 paraffin-embedded specimens of different histological origin. C/T TAA appears to be expressed in a high percentage (>50%) of cancer cells from different tumor types such as lung,
skin, gynecological, stomach and gall bladder cancers. The future characterization of MAGE-A10-specific antibodies might set the stage for the development of targeted active immunotherapy by clarifying potential indications and by allowing the selection of patients eligible for treatment and monitoring of its effectiveness.
4. GENERAL INTRODUCTION

4.1 Cancer

Cancer is a leading cause of death worldwide, according to the International Agency for Research on Cancer (IARC) data, there were 12.7 million new cancer cases in 2008 worldwide, of which 5.6 million occurred in economically developed countries and 7.1 million in economically developing countries. The corresponding estimates for total cancer deaths in 2008 were 7.6 million (about 21,000 cancer deaths a day), 2.8 million in economically developed countries and 4.8 million in economically developing countries [1].

Cancer arises from a loss of normal growth control. In normal healthy tissue, the ratio between cell growth and cell death are in balance. In cancer, this balance is disrupted. This disturbance can result from uncontrolled cell growth or loss of a cell's ability to undergo apoptosis "cell suicide". This is also referred to as programmed cell death and the mechanism by which old or damaged cells can be removed. Cancer being a genetic disease, the genetic material (DNA) of a cell can get damaged, acquiring mutations that affect normal cell growth, division and apoptosis. When this occurs, cells may start to grow uncontrollably and eventually form a mass of tissue called a tumor [2].

Cancer is a disease in which cells display uncontrolled growth, invade and destroy adjacent tissues, and often metastasize to other regions of the body via the lymphatic system or through the bloodstream [2]. These properties discriminate benign from malignant tumors. There are more than 100 different types of cancer. The name of the tumors often comes from the organs or the type of cells affected and can be broadly grouped into the following categories [2]:

a. Carcinoma - a malignant epithelial neoplasm that tends to invade surrounding tissue and to metastasize to distant regions of the body. Carcinomas develop most frequently in the skin, large intestine, lungs, stomach, prostate, cervix, or breast. The tumor is firm, irregular, and nodular, with a well-defined border.
b. **Sarcoma** - a tumor, often highly malignant, composed of cells derived from connective tissue such as bone, cartilage, muscle, blood vessel, or lymphoid tissue.

Sarcomas usually develop rapidly and metastasize through the lymph channels. Different types are named for the specific tissue they affect: **Fibrosarcoma** in fibrous connective tissue; **Lymphosarcoma** in lymphoid tissues; **Osteosarcoma** in bone; **Chondrosarcoma** in cartilage; **Rhabdomyosarcoma** in muscle; and **Liposarcoma** in fat cells.

c. **Leukemia** - a progressive, malignant neoplasm of the blood-forming organs, marked by diffuse replacement of the bone marrow leukocytes and their precursors in the blood and bone marrow, based on cell type can be divided in to myeloid (acute and chronic) and lymphoid (acute and chronic). It is accompanied by a reduced number of erythrocytes and blood platelets, resulting in anemia and increased susceptibility to infection and hemorrhage.

d. **Myeloma** - an osteolytic neoplasm consisting of a profusion of cells typical of the bone marrow that may develop in many sites and cause extensive destruction of the bone. The tumor occurs most frequently in the ribs, vertebrae, pelvic bones, and flat bones of the skull.

e. **Lymphoma** - Cancer that begins in cells of the immune system that affects lymph cells and tissues, including certain white blood cells (T and B cells), lymph nodes, bone marrow, and the spleen. There are two basic categories of lymphomas. One kind is Hodgkin lymphoma, which is marked by the presence of a type of cell called the Reed-Sternberg cell. The other category is non-Hodgkin lymphomas, which develop from lymphocytes — a type of white blood cell. Non-Hodgkin lymphomas can be further divided into cancers that have an indolent (slow-growing) course and those that have an aggressive (fast-growing) course.

f. **Central nervous system cancers** - a neoplasm of the brain or spinal cord that characteristically does not spread beyond the cerebrospinal axis, although it may be highly invasive locally and have widespread effects on body functions. Intracranial neoplasms are about four times more common than those arising in the spinal cord. From 20% to 40% of brain tumors are
metastatic lesions from primary cancer elsewhere, such as in the breast, lung, GI tract, kidney, or a site of melanoma [2].

Cancer can be treated by chemotherapy, surgery, radiation therapy, immunotherapy, monoclonal antibody therapy or other methods (experimental cancer treatments are also under development). The choice of therapy depends on the cancer type, the degree and location of the cancer and the patient's general medical condition [2].
4.2 Cancer Genetics

Cancers arise as consequence of accumulating multiple mutations in the genome. Mutations in genes controlling cell growth are responsible for many major human cancers. Two broad classes of genes have been defined: proto-oncogenes (Table 1) and tumor suppressor genes like Rb, VHL, APC, PTEN, TP53 CD95, ST5, YPEL3, ST7, and ST14 play a key role in cancer initiation and progression. These genes encode many kinds of proteins that help control cell growth and proliferation; mutations in these genes can thus contribute to the development of cancer. Under normal conditions the activities of these two gene classes are optimally balanced. In general, proto-oncogenes allow cell growth whereas tumor suppressor genes inhibit it.

Proto-oncogenes are genes that encode proteins, which can be mutated into a cancer-promoting oncogenes, either by changing the protein-coding segment or by altering its expression. This activation results in a gain of function and drives cell multiplication and thus proliferation.

In contrast tumor suppressor genes encode proteins directly or indirectly inhibiting initiation and/or progression through the cell cycle and other tumor related pathways (e.g. WNT) in which a loss-of-function mutation is oncogenic. A germline mutation in a tumor suppressor gene (e.g. RB, APC, and BRCA1) greatly increases the risk for developing certain types of (hereditary) cancer.

There are also some genetic alterations in DNA that lead to cancer development such as chromosomal translocations and gene promoter (hyper- and hypo-) methylation.

Translocations generate novel chromosomes. In a translocation, a segment from one chromosome is transferred to a nonhomologous chromosome or to a new site on the same chromosome. Translocation events may disrupt gene loci, cause (micro)deletions or result in genes fusion. Some examples are: Ewing’s sarcoma, which is the result of a translocation between chromosomes 11 and 22, which fuses the EWSR1 gene of chromosome 22 to the FLI1 gene of chromosome 11; chronic myelogenous leukemia (CML) which is the result of a reciprocal translocation between chromosome 9 and 22 (Philadelphia chromosome).

DNA methylation is an important regulator of gene transcription, and its role in
carcinogenesis has been a topic of considerable interest in the last few years. Alterations in DNA methylation are common in a variety of tumors as well as in development. Of all epigenetic modifications, hypermethylation, which represses transcription of the promoter regions of tumor suppressor genes leading to gene silencing, has been most extensively studied [3].

DNA methylation mainly consists of a covalent chemical modification, resulting in the addition of a methyl (CH3) group at the carbon 5 position of the cytosine ring. Even though most cytosine methylation occurs in the sequence context 5′CG3′ (also called the CpG dinucleotide), some involves CpA and CpT dinucleotides [4, 5]. Numerous genes have been found to undergo hypermethylation in cancer like genes involved in cell cycle regulation ($p16^{INK4a}$, $p15^{INK4a}$, Rb, $p^{14ARF}$) genes associated with DNA repair (BRCA1, MLH1, MGMT), apoptosis (DAPK, TMS1), drug resistance, detoxification, differentiation, angiogenesis, and metastasis [5]. Although certain genes such as RASSF1A and $p16$ are commonly methylated in a variety of cancers, other genes are methylated in specific cancers. One example is the GSTP1 gene, which is hypermethylated in more than 90% of prostate cancers but is largely unmethylated in acute myeloid leukemia [5]. DNA hypomethylation, on the other hand, is observed in a wide variety of malignancies [6, 7]. It is common in solid tumors such as metastatic hepatocellular cancer and prostate tumors [8, 9].
Table 1: Oncogenes commonly altered in human cancer.

<table>
<thead>
<tr>
<th>Oncogene</th>
<th>Function</th>
<th>Alteration in Cancer</th>
<th>Neoplasm</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKT1</td>
<td>Serine/threonine</td>
<td>Amplification</td>
<td>Gastric carcinoma</td>
</tr>
<tr>
<td></td>
<td>kinase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKT2</td>
<td>Serine/threonine</td>
<td>Amplification</td>
<td>Ovarian, breast, pancreas cancer</td>
</tr>
<tr>
<td></td>
<td>kinase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRAF</td>
<td>Serine/threonine</td>
<td>Point mutation</td>
<td>Melanoma, lung, colorectal cancer</td>
</tr>
<tr>
<td></td>
<td>kinase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTNNB1</td>
<td>Signal transduction</td>
<td>Point mutation</td>
<td>Colon, prostate, melanoma, skin, others</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOS</td>
<td>Transcription fact</td>
<td>Overexpression</td>
<td>Osteosarcoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERBB2</td>
<td>Receptor tyrosine</td>
<td>Point mutation, amplification</td>
<td>Breast, ovary, stomach, neuroblastoma</td>
</tr>
<tr>
<td></td>
<td>kinase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JUN</td>
<td>Transcription fact</td>
<td>Overexpression</td>
<td>Lung cancer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>Receptor tyrosine</td>
<td>Point mutation, rearrangement</td>
<td>Osteosarcoma, kidney, glioma</td>
</tr>
<tr>
<td></td>
<td>kinase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYB</td>
<td>Transcription fact</td>
<td>Amplification</td>
<td>AML, CML, colon cancer, melanoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-MYC</td>
<td>Transcription fact</td>
<td>Amplification</td>
<td>Breast, colon, gastric, lung</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-MYC</td>
<td>Transcription fact</td>
<td>Amplification</td>
<td>Lung carcinoma, bladder</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-MYC</td>
<td>Transcription fact</td>
<td>Amplification</td>
<td>Neuroblastoma, lung cancer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRAS</td>
<td>GTPase</td>
<td>Point mutation</td>
<td>Colon, lung, pancreas</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KRAS</td>
<td>GTPase</td>
<td>Point mutation</td>
<td>Melanoma, colorectal cancer, AML</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRAS</td>
<td>GTPase</td>
<td>Point mutation</td>
<td>Various carcinomas, melanoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REL</td>
<td>Transcription fact</td>
<td>Rearrangement, amplification</td>
<td>Lymphoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WNT1</td>
<td>Growth factor</td>
<td>Point mutation</td>
<td>Retinoblastoma</td>
</tr>
</tbody>
</table>

Compilation of oncogenes commonly altered in human cancer, taken from [2]:
AML, acute myelogenous leukemia; CML, chronic myelogenous leukemia.
4.2.1 Colorectal Cancer

The colon is part of the digestive system, which is a series of bodily organs beginning at the mouth and ending with the anus, and it is responsible for the final stages of the digestive process (Figure 1). The colon's function is threefold: to absorb the remaining water and electrolytes from indigestible food matter; to accept and stores food remains that were not digested in the small intestine; and to eliminate solid waste (feces) from the body. The rectum is the segment from end of the colon to the anus. Together, they form a long, muscular tube called the large intestine (also known as the large bowel). Tumors of the colon and rectum are growths arising from the inner wall of the large intestine.

Precursor lesions of the large intestine are called polyps and has been defined as a benign neoplastic growth with variable malignant potential, representing proliferation of epithelial tissue in the lumen of the sigmoid colon, rectum, or stomach, polyps do not invade nearby tissue or spread to other parts of the body. The majority of colorectal cancer (CRC) arises from malignant transformation of an adenomatous polyp (Figure 2).

CRC is the fourth most common cancer in men and the third most common cancer in women worldwide [10, 11]. The American Cancer Society's most recent estimates for the number of colorectal cancer cases in the United States for 2011 are: 101,700 new cases of colon cancer and 39,510 of rectal cancers with 49,380 deaths.

Risk factors according to the American Cancer Society for the development of CRC are:

a. **Age**: peak incidence in the 6\textsuperscript{th}/7\textsuperscript{th} decade.

b. **Nutrition**: a diet that is high in red meat, processed meats and not enough fruits and vegetables consumption is a colon cancer risk.

c. **A family history of colorectal cancer**: A family history of colorectal cancer or adenomatous polyps increases the risk of colorectal cancer. That risk is even higher if the close relative was diagnosed at a young age or if more than one 1\textsuperscript{st} degree family member had colon cancer.
d. **Member of certain racial or ethnic groups**: African Americans get colon cancer more often than other racial groups in the U.S. and are nearly twice as likely to die from it.

e. **Inherited conditions such as familial adenomatous polyposis**, which causes the development of 100-1000 of polyps in the colon, also raises the risk of colorectal cancer.

f. **Inflammatory bowel disease**: ulcerative colitis and Crohn's disease are a bowel disease characterized by inflammation with ulcer formation in the lining of the colon (large intestine). People with either of these diseases can develop expansion of immature cells, with a corresponding decrease in the number and location of mature cells, this is called dysplasia, which may turn into cancer [2].

**Figure 1: Anatomy of the colon.**

![Anatomy of the colon](image)

Depiction of colorectal anatomy, taken from [12].

A: anatomy of the colon; B: anatomy of the rectum.

CRC progresses through a series of clinical and histopathological stages ranging from single crypt lesions through small benign tumors (adenomatous
polyps) to malignant cancers (carcinomas) (Figure 2). Stages are usually defined by TNM classification, where T describes the size of the tumor and whether it has invaded nearby tissue, N describes regional lymph nodes that are involved, M describes distant metastasis (spread of cancer from one body part to another). The depth of tumor invasion defines the T stage and increases from T1 (invasion of the submucosa) to T4 (invasion into the serosa or adjacent structures) [13]. Another grading system is Dukes classification that considers the arrangement of the cells rather than the percentage of the differentiated cells. The initial Dukes approach has evolved into the three-grade system. Grade 1 is the most differentiated, with well-formed tubules and the least nuclear polymorphisms and mitoses. Grade 3 is the least differentiated, with only occasional glandular structures, pleomorphic cells and a high incidence of mitoses. Grade 2 is intermediate between Grades 1 and 3 [14].

The development of colorectal cancer is a multistep process that involves an accumulation of mutations in tumor suppressor genes and oncogenes. The model of colorectal tumorigenesis includes several genetic changes that are required for cancer initiation and progression [15]. The earliest trigger genetic event is the inactivation of the APC (adenomatous polyposis coli) gene (WNT pathway). Mutations in other tumor suppressor genes like SMAD4 and TP53 and oncogenes like KRAS and likely several other genes/pathways accompany transitions (in the pathology) of the lesions and drive tumors towards malignancy and metastasis [16]. Alongside with gene mutations, deregulated expression of oncogenes and/or tumor suppressor genes can also occur following epigenetic modifications of their promoters. In chapter 6.2 we describe SH2D4A gene as novel tumour suppressor in colorectal cancer, that physically interacts with the EGFR/STAT3 pathway and controls cell proliferation. Upon EGF signalling, Shoca-2 recruits the serine/threonine phosphatase PP1β to the receptor complex and represses activated STAT3 via dephosphorylation. Shoca-2 expression reduces anchorage-independent tumour cell growth and its loss promotes the expression of c-Myc, Cyclin D1 and Jun B. Shoca-2 expression may be lost in human colorectal cancers as a result of chromosomal instability, mutations and epigenetic changes and
diminished Shoca-2 protein expression correlates with advanced disease stages and is associated with poor prognosis.

Figure 2: Stages of colon cancer.

The diagram, taken from [2], above illustrates the progression of polyp to a cancer and the cancer's subsequent progression if left untreated.

4.3 Hereditary cancer syndromes
The majority of cancers occur without any predisposition, although genetic factors are believed to have an impact on the susceptibility of an individual to develop cancer [17]. Approximately 5-10% of all cancers are hereditary, which means that germline mutations in specific genes that are known to be related to cancer development can be passed on to offspring. Individuals who inherit one of these gene changes will have a higher likelihood of developing cancer within their lifetime. Inherited cancer syndromes with their genetic causes are listed in Table 2.

**Table 2: Examples of hereditary cancer syndromes.**

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Gene(s)</th>
<th>Location(s)</th>
<th>Tumor types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast/Ovarian</td>
<td>BRCA1</td>
<td>17q21</td>
<td>Breast, ovarian</td>
</tr>
<tr>
<td>Breast/Ovarian</td>
<td>BRCA2</td>
<td>13q12.3</td>
<td>Breast, ovarian, prostate</td>
</tr>
<tr>
<td>Cowden syndrome</td>
<td>PTEN</td>
<td>10q23.3</td>
<td>Pituitary, testicle, thyroid, breast</td>
</tr>
<tr>
<td>Familial Adenomatous Polyposis</td>
<td>APC</td>
<td>5q21</td>
<td>GI tract, thyroid</td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>MLH1, MSH2, MSH6, PMS2</td>
<td>2p22-21, 3p21, 2p16, 2q31-33, 7p22</td>
<td>GI tract, ovary, endometrium</td>
</tr>
<tr>
<td>Juvenile polyposis coli</td>
<td>SMAD4, BMP1A</td>
<td>18q21.1, 10q23.2</td>
<td>GI tract</td>
</tr>
<tr>
<td>Li-Fraumeni</td>
<td>TP53</td>
<td>17p13</td>
<td>Breast, soft tissue, brain, leukaemia</td>
</tr>
<tr>
<td>Neurofibromatosis 1</td>
<td>NF1</td>
<td>17q11</td>
<td>Nervous system, skin, muscle, leukaemia</td>
</tr>
<tr>
<td>Neurofibromatosis 2</td>
<td>NF2</td>
<td>22q12</td>
<td>Schwann cells, spine skin</td>
</tr>
<tr>
<td>Peutz-Jeghers</td>
<td>STK1</td>
<td>19p13.3</td>
<td>Gastointestinal tract</td>
</tr>
</tbody>
</table>

In hereditary cancer predisposing syndromes, the critical gene is mutated only on one allele but no on the second, “wild type” one. According to Knudson’s widely accepted “two hit theory” which was based on the epidemiological studies of retinoblastoma gene (Rb) [18], both of the alleles of a tumor suppressor gene need to be inactivated before tumorigenesis ensues. Thus, the germline mutation alone is not sufficient to give rise to tumor formation, but the functional wild type allele needs to be inactivated before neoplastic reactions may occur [19].
There are several known and proposed mechanisms, which can cause the inactivation of the wild type allele. It can occur for example by loss of heterozygosity (LOH), somatic deletion [20], or epigenetically by promoter methylation [21]. However, to explain the sometimes even 100% penetrance of hereditary cancer syndromes resulting from a heterozygous mutation in a tumor suppressor gene, it is nowadays believed that also “recessive” mutation have an effect on the tissue phenotype and thus result in a clonal expansion of mutated cells [22]. Dominantly functioning oncogenes can accelerate tumorigenesis also when heterozygous.

4.3.1 Lynch syndrome (aka HNPCC)

Lynch syndrome (LS) represents the most common, autosomal dominantly
inherited cancer predisposition worldwide with incomplete penetrance, accounting for approximately 1-6% of all colorectal malignancies. It is characterized by an increased lifetime risk for colon cancer and cancers of the endometrium, ovary, stomach, small intestine, hepatobiliary tract, urinary tract, brain, and skin [23-25].

It is caused by a germline mutation in a mismatch repair gene (MLH1, MSH2, PMS2, and MSH6), MSH2 on chromosome 2p, and MLH1 on chromosome 3p account for the majority of genetically defined cases. 50-60% of families with the clinical diagnosis of Lynch syndrome are found to have mutations in these genes; the mutation prevalence depends on features of the family history [26, 27].

The number of polyps is modest without endoscopical differences compared to sporadic polyps but with an early onset (44 years), bigger size and frequency, and with more villous and dysplastic features as well as a more rapid progression to cancer [28, 29]. Without intervention, patients with Lynch syndrome have an 80% lifetime risk of colorectal cancer [30] however the risk of cancer-related mortality remains considerable. Nevertheless, patients with Lynch syndrome-related colorectal cancer appear to have better survival rates when compared to patients with sporadic colorectal cancer [31]. Tumors in LS patients characteristically exhibit microsatellite instability (MSI). The diagnostic settings for LS diagnosis are based on Amsterdam criteria (constantly update) to identify the affected family and the Bethesda guidelines to define the tumors (Table 2) [32-37]. Essentially this guideline specifies that at least three relatives be affected with colorectal cancer, endometrial cancer (female patients), possibly other LS-related malignancies and MSI-H. Genetic testing is improving the approach to patients at risk for LS. The American Gastroenterological Association recommends a testing strategy beginning with MSI testing on tumor tissue from individuals satisfying the revised Bethesda criteria. In persons with MSI-high tumors, germline analysis of hMSH2, hMSH1 and hMSH6 should be assessed. For families in which tumor tissue is not available, germline testing is recommended if any of the first three Bethesda criteria are met (Table 2) [38]. Failure to detect a clearly pathogenic mutation in the proband doesn’t exclude Lynch syndrome.

Immunohistochemistry (IHC) testing for the expression of mismatch related
proteins in tumor tissue is a complementary screening method [39], The MMR gene products work in heterodimers: MSH2 with MSH6 or MSH3 protein, and MLH1 with PMS2 or PMS1 protein. A germline mutation in *MSH2* typically results in loss of expression of both proteins MSH2 and MSH6 and a germline mutation in *MLH1* in loss of expression of the proteins MLH1 and PMS2. Germline mutations in *MSH6* and *PMS2* typically however do not result in loss of MSH2 or MLH1 expression because these proteins are still present in other pairings and therefore remain stable expressed [40].

For patients with a definitive or suspected diagnosis of LS, interval surveillance is recommended encompassing: full colonoscopy every one to two years beginning between the ages of 20 and 25 and endometrial and ovarian cancer screening beginning between the of 25 and 35 (for female) [41, 42].

In appendix 8.2 analyzing 12 samples from 6 LS patients (CRC and matched cancer free mucosa) with Affymetrix Whole Genome 2.7M chip and validating the results in a large cohort of LS-related CRCs (50 patients), we describe 3 new hotspot frequently deleted or duplicated in LS related colorectal cancers.

<table>
<thead>
<tr>
<th>Clinical criteria for LS.</th>
</tr>
</thead>
</table>

17
<table>
<thead>
<tr>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amsterdam</strong></td>
</tr>
<tr>
<td>1. Three or more relatives with colorectal cancer, one of whom is a first-degree relative of the other two;</td>
</tr>
<tr>
<td>2. Colorectal cancer involving at least two generations;</td>
</tr>
<tr>
<td>3. One or more colorectal cancers diagnosed at age &lt;50 years</td>
</tr>
<tr>
<td><strong>Amsterdam II</strong></td>
</tr>
<tr>
<td>1. Three or more relatives with a HNPCC-associated cancer (colorectal, endometrial, small bowel, ureter, renal pelvis), one of whom is a first-degree relative of the other two;</td>
</tr>
<tr>
<td>2. Colorectal cancer involving at least two generations;</td>
</tr>
<tr>
<td>3. One or more colorectal cancers diagnosed at age &lt;50 years</td>
</tr>
<tr>
<td><strong>Modified Amsterdam</strong></td>
</tr>
<tr>
<td>1. Very small families which cannot be further expanded can be considered as HNPCC with only two colorectal cancers in first-degree relatives; colorectal cancer must involve at least two generations, and at least one colorectal cancer must be diagnosed at age &lt;55 years</td>
</tr>
<tr>
<td>2. In families with two first-degree relatives affected by colorectal cancer, the presence of a third relative with an unusual early-onset neoplasm or endometrial cancer is sufficient</td>
</tr>
<tr>
<td><strong>Young age at onset</strong></td>
</tr>
<tr>
<td>Young age at onset Proband diagnosed at age &lt;40 years without a family history fulfilling Amsterdam or Modified Amsterdam criteria</td>
</tr>
<tr>
<td><strong>HNPCC-variant</strong></td>
</tr>
<tr>
<td>Family history suggestive of HNPCC, but not fulfilling Amsterdam, Modified Amsterdam, or young age at onset criteria</td>
</tr>
<tr>
<td><strong>Bethesda</strong></td>
</tr>
<tr>
<td>1. Amsterdam criteria</td>
</tr>
<tr>
<td>2. Two HNPCC-related cancers, including synchronous and metachronous colorectal cancers or associated extracolonic cancers</td>
</tr>
<tr>
<td>3. Colorectal cancer and a first-degree relative with colorectal cancer and/or a HNPCC-related extracolonic cancer and/or a colorectal adenoma; one of the cancers diagnosed at age &lt;45 years, and the adenoma diagnosed age &lt;40 years</td>
</tr>
<tr>
<td>4. Colorectal cancer or endometrial cancer diagnosed at age &lt;45 years</td>
</tr>
<tr>
<td>5. Right-sided colorectal cancer with an undifferentiated pattern on histology diagnosed at age &lt; 45 years</td>
</tr>
<tr>
<td>6. Signet-ring cell-type colorectal cancer diagnosed at age &lt;45 years</td>
</tr>
<tr>
<td>7. Adenomas diagnosed at age &lt;40 years</td>
</tr>
<tr>
<td><strong>Revised Bethesda</strong></td>
</tr>
<tr>
<td>1. Colorectal cancer diagnosed at age &lt;50 years</td>
</tr>
<tr>
<td>2. Synchronous or metachronous colorectal or other HNPCC-associated tumors regardless of age</td>
</tr>
<tr>
<td>3. Colorectal cancer diagnosed at age &lt;60 years with histologic findings of infiltrating lymphocytes, Crohn’s-like lymphocytic reaction, mucinous/signet ring differentiation or medullary growth pattern.</td>
</tr>
<tr>
<td>4. Colorectal cancer in $\geq$1 first-degree relative(s) with an HNPCC-related tumor, with one of the cancers being diagnosed at age &lt;50 years</td>
</tr>
<tr>
<td>5. Colorectal cancer diagnosed in $\geq$2 first- or second-degree relatives with HNPCC-related tumors, regardless of age</td>
</tr>
</tbody>
</table>

Table taken from [43]
4.3.1.1 Mismatch repair system

The DNA Mismatch repair system has an important role in maintaining genomic stability during DNA replication by correcting mismatches and small insertion or deletion loops (IDLs) introduced through errors made by DNA polymerases. Components of the MMR system MutS, MutL, MutH and Uvr were identified in Escherichia coli through the genetic studies of mutants [44, 45] and the system has been well conserved during the evolutionary process in all eukaryotic organisms (yeast, mouse and human), [46, 47].

DNA mismatch repair (MMR) can be divided in three phases: initiation, excision and resynthesis (Figure 3). The three MutS homologs, MSH2, MSH3 and MSH6 are involved in the initiation of MMR. The MutS homologs forming heterodimers, recognize DNA damage; the MSH2 and MSH6 dimer (the MutSα complex) recognizes mismatches base-base and single base loops while the MSH2 and MSH3 dimer (hMutSβ complex) identifies deletion/insertion loops of more than one base [48-50].

After recognition of the DNA damage the heterodimer MutLα, formed by MLH1 and PMS2, binds MutSα and mediates the interaction between the MutS proteins and enzymes involved in long-patch excision in postreplication mismatch repair [51].

Other proteins are involved in the excision of the damaged strand and resynthesis steps of MMR that are recruited subsequently. Proteins known to be involved are proliferating cell nuclear antigen (PCNA) that is known to work as a processivity factor for replicative polymerases [52], with interactions with MSH2 and MLH1 [53], and with MSH6 [54], as well as exonuclease Exol, and DNA polymerase δ I [55, 56].
Figure 3: The mismatch repair system (MMR).

A. Initiation

B. Excision & resynthesis

The mismatch repair system (MMR), taken from Helleman et al. [57].
A. Initiation of MMR by recognizing the DNA damage by the MutSα or β complex and recruiting the MutLα complex. B. Excision of the damaged strand and resynthesis in which exonuclease Exol, proliferating cell nuclear antigen (PCNA), DNA polymerase δ or ε and DNA helicase I are suggested to play a role.
4.3.1.2 Microsatellite instability

The inactivation of the MMR system leads to widespread genomic instability, because the DNA-replication errors induced DNA polymerase slippage during replication of long repetitive DNA sequences. As a consequence, these tumors (MMR deficients) exhibit a specific phenotype called microsatellite instability (MSI), affecting mono-, di-, tri- and tetra nucleotide repetitive sequences [58, 59].

MSI in tumors can be initiated by genetic or epigenetic inactivation of MMR genes. Mouse knockout models have demonstrated that MLH1<sup>-/-</sup>, PMS2<sup>-/-</sup> MSH2<sup>-/-</sup> and MSH3<sup>-/-</sup>, exhibit a high frequency of MSI in tumors while MSH6<sup>-/-</sup> and PMS1<sup>-/-</sup> mice a had a low frequency [60].

Five markers (two mononucleotide repeats: BAT25 and BAT26 and three dinucleotide repeats: D2S123, D5S346, and D17S250) have been recommended by the National Cancer Institute to screen for MSI in LS-related CRCs (often called Bethesda markers) [61]. Tumors with instability at two or more of these markers were defined as MSI-High (MSI-H), while those with instability at one repeat or display no instability were defined as MSI-Low (MSI-L) and those without as microsatellite stable (MSS) tumors, respectively [62]. They occur usually in the Lynch syndrome or sporadically in as many as 10–15% of colorectal, gastric, and endometrial (20-30%) carcinomas [63-65]. Repetitive segments of DNA (dinucleotide, trinucleotide, tetranucleotide, pentanucleotide repeats), scattered throughout the genome in non-coding regions between genes or within genes (introns), often used as markers for linkage analysis because of the naturally occurring high variability in repeat number between individuals. These regions are inherently unstable and susceptible to mutation, Sporadically these repetitive sequences are located in the coding region of genes. The TGFβRII, BAX, IGFIIR, MSH3, and MSH6 genes were the first described genes for instability in MSI-H cancers [66-69]. More recently, additional genes with similar mutations were described encompassing ACTRII, AIM2, APAF-1, AXIN-2, BCL-10, BLM, Caspase-5, CDX-2, CHK-1, FAS, GRB-14, cell cycle protein hG4-1, KIAA0977, MBD-4, hMLH3, NADH ubiquinone oxidoreductase, OGT, PTEN, RAD-50, RHAMM, RIZ, SEC63, SLC23AT, TCF-4, and WISP-3. MSI-H CRCs [59, 62, 70-84],
whether hereditary or sporadic (through somatic MLH1 promoter hypermethylation), are generally associated with a more favorable prognosis when compared to MSI-L and MSS ones [85]. Recent genome-wide gene expression data on (sporadic) MSS and MSI-H CRCs further demonstrate that tumor development in microsatellite unstable, MMR deficient cancers follows distinct pathogenetic alleys. In chapter 6.1 we identified a mononucleotide repeat tract (EWS16T) within the 3'UTR of the EWSR1 gene that is consistently mutated in cancers with microsatellite instability and thus represents a novel target gene locus in mismatch repair-deficient tumors. Our data indicate that contractions at this locus promote SFPQ-mediated distal poly(A) site usage in EWSR1 pre-mRNAs resulting in decreased EWS expression that is accompanied by altered subcellular localization of the protein.
4.4 Other hereditary colorectal cancer syndromes

4.4.1 Familial adenomatous polyposis

Familial adenomatous polyposis (FAP) is a an autosomal dominantly inherited disorder with 100% penetrance by age 40 years characterized by the presence of more than hundreds/thousands of benign polyps (adenomas) some of which, inevitably progress to colorectal cancer in the third and fourth decade of life [86, 87]. In addition an extracolonic cancer spectrum may occur, including polyps of the upper gastrointestinal tract, desmoids, osteomas, fibromas, and congenital hypertrophy of the retinal pigment epithelium (CHRPE) [88]. It is thought to be responsible for about 0.5-1% of colorectal cancers diagnosed each year with an estimated prevalence of approximately 1/10,000 [89]. It is caused by germline mutations in the APC (adenomatous polyposis coli) gene located on the long arm of chromosome 5 (5q21). The majority of mutations (>90%) are nonsense or frameshift mutations that lead to premature stop codons. The resulting protein is truncated and apparently nonfunctional. The bi-allelic inactivation (1st germline and 2nd somatic hit) of the APC have a direct role to the disruption of Wnt signalling pathway. Indeed, it is well understood that APC acts as a down regulator of β-catenin; and the mutant APC protein would therefore lead to the accumulation of cytoplasmic β-catenin. This subsequently leads to an increased level of β-catenin/Tcf complex which alters the expression of genes such as c-myc and cyclin D, and then initiating uncontrollable cell growth.

An APC gene mutation can be identified in 80–90% of kindreds with the classical disorder [90]. In addition somatic mutations of the APC gene are found in as many as 80% of sporadic tumors, and seems to be involved in initiating tumorigenesis [91].
4.4.1.1 Familial adenomatous polyposis variants

Attenuated Adenomatous Polyposis Coli (AAPC), Gardner syndrome and Turcot syndrome are the three variants of the FAP syndrome. Extraintestinal manifestations distinguish the first two of these variants. AAPC as the name implies, represents a less pronounced FAP. In addition, biallelic mutations in the \textit{MUTYH} gene have been found in 30\% of families with multiple adenomas (15–100) who do not exhibit an autosomal dominant inheritance pattern or a germline mutation in the \textit{APC} gene [92]. Patients with attenuated adenomatous polyposis coli develop fewer (10-99) polyps than the typical FAP and at a later age. The polyps tend to be distributed more proximally, and the risk of colon adenocarcinoma, appear to be lower than in a classical FAP[93]. Beside polyposis Gardner syndrome includes the development of benign extracolonic tumors, desmoid tumors, soft-tissue tumors, osteomas and dental abnormalities. Although extremely rare in the general population, desmoids occur in approximately 10\% of patients with FAP, most often in the mesentery or abdominal wall following surgical trauma. Despite they do not metastasize, they are locally aggressive and a common cause of death in patients with FAP. The distinction between Gardner syndrome and FAP has become less clear with improved understanding of the genetic basis of these disorders. The same genetic abnormalities are found in patients with either syndrome and both syndromes can be seen within the same family [94]. The additional feature of Turcot syndrome is the growth of central nervous system tumors like medulloblastomas, astrocytomas and ependymomas. In about 70-80\% of classical FAP an \textit{APC} gene mutation can be detected. Nevertheless, in few patients with Turcot syndrome, \textit{MLH1} and \textit{MSH2} genes mutation have been described [95].
4.4.2 Hamartomatous polyposis syndromes

This group of disorders has in common the development of hamartomatous polyps in the GI tract, which may also confers an increased risk of cancer. These conditions include Juvenile Polyposis Syndrome (JPS), Peutz-Jeghers Syndrome (PJS), Bannayan-Riley-Ruvalcaba Syndrome (BRRS), Cowden Syndrome (CS), Cronkhite-Canada Syndrome (CCS), and Hereditary Mixed Polyposis Syndrome (HMPS).

4.4.2.1 Juvenile polyposis

JPS is characterized by multiple distinct juvenile polyps in the gastrointestinal tract and an increased risk of colorectal cancer. Clinically it is defined by the presence of five or more juvenile polyps in the colorectum, juvenile polyps throughout the gastrointestinal tract or any number of juvenile polyps, and a positive family history of juvenile polyposis [96, 97].

Juvenile polyposis syndrome (JPS) is a rare (2.8 per 100 000 in children under 10 years of age) autosomal dominant inherited disorder with germline mutations in the SMAD4 or BMPR1A gene in about 50%-60% of JPS patients [98, 99], and about 10% of JPS patients with a germline mutation in PTEN [100].

Patients present with bleeding, anemia, rectal prolapse of a polyp, diarrhea, abdominal pain and/or failure to thrive and in a small subset of patients congenital abnormalities, including cardiac, and bowel rotations have been shown. The most significant cause of mortality in patients with juvenile polyposis is colorectal cancer. In one large series, 18 of 87 patients (21%) developed colorectal cancer. These cancers tended to have early onset (mean age 34 years), were poorly differentiated and had a poor prognosis [96]. Overall, the risk of gastrointestinal malignancy may exceed 50% [101].
4.4.2.2 Peutz–Jeghers syndrome

Peutz–Jeghers syndrome (PJS) is a rare autosomal dominant condition (1 in 150,000 persons), characterized by the development of hamartomatous polyps, benign polyps comprised of several types of typical epithelial cells supported by a thick band of smooth muscle. They are found throughout the gastrointestinal tract (particularly the small bowel) and may cause bleeding, intussusception, or obstruction [102]. Patients usually develop distinctive mucocutaneous pigmented lesion on lips, buccal mucosa, hands, and feet. The lifetime risk of colon cancer is about 39%, and 93% for stomach, small bowel, pancreas, breast, sex cord, uterine, cervical, and melanoma cancers [103-107].
PJS has been associated with germline mutations or deletions in LKB1 (STK11), a serine–threonine kinase that regulates p53-mediated apoptosis and the mammalian target of rapamycin (mTOR) pathway; STK11 germline alterations can be identified in only 50%-60% of cases of suspected PJS [108-110].
4.4.2.3 PTEN Hamartoma Tumor Syndrome

The *PTEN* hamartoma tumor syndrome (PHTS) includes Cowden syndrome (CS), Bannayan-Riley-Ruvalcaba syndrome (BRRS), *PTEN*-related Proteus syndrome (PS), and Proteus-like syndrome are rare polyposis syndromes related to juvenile polyposis. CS is a multiple hamartoma syndrome with a high risk for benign and malignant tumors of the thyroid, breast, and endometrium [111]. Affected individuals usually have macrocephaly, trichilemmomas, and papillomatous papules and present by the late 20s. The lifetime risk of developing breast cancer is 25%-50%, with an average age of diagnosis between 38 and 46 years. The lifetime risk for thyroid cancer (usually follicular, rarely papillary, but never medullary thyroid cancer) is approximately 10%. The risk for endometrial cancer, although not well defined, may approach 5%-10% [111]. BRRS is a congenital disorder characterized by hamartomatous polyps as well as unusual facies, macrocephaly, developmental delay, and pigmented papules on the penis [43]. PS is a complex, highly variable disorder involving congenital malformations and hamartomatous overgrowth of multiple tissues, as well as connective tissue nevi, epidermal nevi, and hyperostoses [111]. Proteus-like syndrome is undefined but refers to individuals with significant clinical features of PS who do not meet the diagnostic criteria for PS [111]. The diagnosis of PHTS is made only when a *PTEN* mutation is identified. Up to 85% of individuals who meet the diagnostic criteria for CS and 65% of individuals with a clinical diagnosis of BRRS have a detectable *PTEN* mutation. Preliminary data also suggest that up to 50% of individuals with a Proteus-like syndrome and up to 20% of individuals with Proteus syndrome have *PTEN* mutations [111].

*PTEN* is a tumor suppressor gene. It acts as a dual-specificity protein phosphatase, dephosphorylating tyrosine-, serine- and threonine-phosphorylated proteins. Also acts as a lipid phosphatase and this activity is critical for its tumor suppressor function. Antagonizes the PI3K-AKT/PKB signaling pathway by dephosphorylating phosphoinositides and thereby modulating cell cycle progression and cell survival. The unphosphorylated
form cooperates with AIP1 to suppress AKT1 activation. Dephosphorylates tyrosine-phosphorylated focal adhesion kinase and inhibits cell migration and integrin-mediated cell spreading and focal adhesion formation. Plays a role as a key modulator of the AKT-mTOR signaling pathway [112].
4.5 Diagnostic and prognostic genetic markers in cancer

Tumor development starts through the accumulation of genetic and epigenetic alterations that allow cells to escape from the cellular control mechanisms that regulate the homeostatic equilibrium between cell proliferation and cell death [113]. To transform a primary cell into a malignant one in vitro requires alterations in the functionality of many processes by which cells control their growth, division and differentiation [114]. Studies on colorectal, breast, lung, pancreatic cancer tissues have demonstrated the view that acquisition of genetic alterations and epigenetic changes could determine the morphological changes that leads to cancer progression [15, 115]. Mutations in cancer cells comprises a large variety of DNA alterations including chromosome copy number, (micro)deletions/(micro)duplications or structural chromosomal alterations like translocations, deletions or amplifications, as well as changes at nucleotide level such as point mutations affecting a single nucleotide at a critical position in a cancer-related gene. It has been shown that these alterations often co-exist in a single tumor [116]. Genetic and epigenetic events represent two complementary mechanisms that are involved in carcinogenesis initiation/progression/metastasis and sometimes coexisting [117].

Biomarkers can be defined as biological molecules found in blood, other body fluids, or tissues, which can indicate a normal or abnormal process, or directly a condition or disease. A biomarker may be further used to see how well the body responds to a treatment for a disease or condition [2, 118]. Biomarkers are an reliable indicators for a disease process and can be divided in five categories: diagnostic, early detection, prognostic and predictive [118]. Each marker should have different characteristics such as specificity, refers to the quantity of control subjects who test negative for the biomarker (patients without disease), and sensitivity refer to quantity of case subject who test positive for the biomarker (patients with confirmed disease).

Some examples are: *TP53* has been found mutated in many tumors and is used as biomarker in epidemiology and for early cancer detection [119-122]; *EGFR* is often mutated in lung adenocarcinomas of never smokers (with
additional mutation in *KRAS*) [123] and is used as biomarker for predicting the response of lung cancer patients to small molecule Tyrosine Kinase Inhibitors (TKIs) such as erlotinib or gefinitib [124-127]; *de novo* methylation of the \( p16^{INK4a} \) promoter is one of the most frequent epigenetic alterations in human cancer and appears to be the earliest event in some cancer types [128, 129]; \( p16^{INKa}, p15^{INK4b}, \text{RASSF1A, MLH1, GSTP1, CDH1, APC, and DAPK1} \) are frequently methylated in circulating DNA and used as biomarkers in the clinic and risk assessment [130]; miR-141 in prostate cancer patients has been shown to be a marker that can distinguish, with significant specificity and sensitivity, patients with cancer from healthy controls [131]. Thus, the discovery of novel gene/pathway alterations could reveal new biomarkers for cancer detection, diagnosis and prognosis.
Tumor markers currently use in clinical practice.

<table>
<thead>
<tr>
<th>Tumor Marker</th>
<th>Cancer Type</th>
<th>Tissue Analyzed</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALK gene rearrangements</td>
<td>Non-small cell lung cancer; anaplastic large cell lymphoma</td>
<td>Tumor tissue</td>
<td>To help determine treatment and prognosis</td>
</tr>
<tr>
<td>Alpha-fetoprotein (AFP)</td>
<td>Liver cancer; germ cell tumors</td>
<td>Blood</td>
<td>To help diagnose liver cancer and follow response to treatment; to assess stage, prognosis, and response to treatment of germ cell tumors</td>
</tr>
<tr>
<td>Beta-2-microglobulin (B2M)</td>
<td>Multiple myeloma; chronic lymphocytic leukemia; some lymphomas</td>
<td>Blood, urine, or cerebrospinal fluid</td>
<td>To determine prognosis and to follow response to treatment</td>
</tr>
<tr>
<td>Beta-human chorionic gonadotropin (Beta-hCG)</td>
<td>Choriocarcinoma; testicular cancer</td>
<td>Urine or blood</td>
<td>To assess stage, prognosis, and response to treatment of testicular cancer</td>
</tr>
<tr>
<td>BCR-ABL</td>
<td>Chronic myeloid leukemia</td>
<td>Blood and/or bone marrow</td>
<td>To confirm diagnosis and monitor disease status</td>
</tr>
<tr>
<td>BRAF mutation V600E</td>
<td>Cutaneous melanoma; colorectal cancer</td>
<td>Tumor tissue</td>
<td>To predict response to targeted therapies</td>
</tr>
<tr>
<td>CA15-3/CA27.29</td>
<td>Breast cancer</td>
<td>Blood</td>
<td>To assess whether treatment is working or disease has recurred</td>
</tr>
<tr>
<td>CA19-9</td>
<td>Pancreatic cancer; gallbladder cancer; bile duct cancer; gastric cancer</td>
<td>Blood</td>
<td>To assess whether treatment is working</td>
</tr>
<tr>
<td>CA-125</td>
<td>Ovarian cancer</td>
<td>Blood</td>
<td>To help in diagnosis, assessment of response to treatment, and evaluation of recurrence</td>
</tr>
<tr>
<td>Calcitonin</td>
<td>Medullary thyroid cancer</td>
<td>Blood</td>
<td>To aid in diagnosis, to check whether treatment is working, and to assess recurrence</td>
</tr>
<tr>
<td>Carcinoembryonic antigen (CEA)</td>
<td>Colorectal cancer; breast cancer</td>
<td>Blood</td>
<td>To check whether colorectal cancer has spread; to look for breast cancer recurrence and assess response to treatment</td>
</tr>
<tr>
<td>CD20</td>
<td>Non-Hodgkin lymphoma</td>
<td>Blood</td>
<td>To determine whether treatment with a targeted therapy is appropriate</td>
</tr>
<tr>
<td>Chromogranin A (CgA)</td>
<td>Neuroendocrine tumors</td>
<td>Blood</td>
<td>To help in diagnosis, assessment of treatment response, and evaluation of recurrence</td>
</tr>
<tr>
<td>Test</td>
<td>Cancer Type</td>
<td>Sample Type</td>
<td>Use</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>------------------------------------</td>
<td>-------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Chromosomes 3, 7, 17, and 9p21</td>
<td>Bladder cancer</td>
<td>Urine</td>
<td>To help in monitoring for tumor recurrence</td>
</tr>
<tr>
<td>Cytokeratin fragments 21-1</td>
<td>Lung cancer</td>
<td>Blood</td>
<td>To help in monitoring for recurrence</td>
</tr>
<tr>
<td>EGFR mutation analysis</td>
<td>Non-small cell lung cancer</td>
<td>Tumor tissue</td>
<td>To help determine treatment and prognosis</td>
</tr>
<tr>
<td>Estrogen receptor (ER)/progesterone receptor (PR)</td>
<td>Breast cancer</td>
<td>Tumor tissue</td>
<td>To determine whether treatment with anti-hormonal therapy (such as tamoxifen) is appropriate</td>
</tr>
<tr>
<td>Fibrin/fibrinogen</td>
<td>Bladder cancer</td>
<td>Urine</td>
<td>To monitor progression and response to treatment</td>
</tr>
<tr>
<td>HE4</td>
<td>Ovarian cancer</td>
<td>Blood</td>
<td>To assess disease progression and monitor for recurrence</td>
</tr>
<tr>
<td>HER2/neu</td>
<td>Breast cancer; gastric cancer; esophageal cancer</td>
<td>Tumor tissue</td>
<td>To determine whether treatment with trastuzumab is appropriate</td>
</tr>
<tr>
<td>Immunoglobulins</td>
<td>Multiple myeloma; Waldenstrom macroglobulinemia</td>
<td>Blood and urine</td>
<td>To help diagnose disease, assess response to treatment, and look for recurrence</td>
</tr>
<tr>
<td>KIT</td>
<td>Gastrointestinal stromal tumor; mucosal melanoma</td>
<td>Tumor tissue</td>
<td>To help in diagnosis and determining treatment</td>
</tr>
<tr>
<td>KRAS mutation analysis</td>
<td>Colorectal cancer; non-small cell lung cancer</td>
<td>Tumor tissue</td>
<td>To determine whether treatment with a particular type of targeted therapy is appropriate</td>
</tr>
<tr>
<td>Lactate dehydrogenase</td>
<td>Germ cell tumors</td>
<td>Blood</td>
<td>To assess stage, prognosis, and response to treatment</td>
</tr>
<tr>
<td>Nuclear matrix protein 22</td>
<td>Bladder cancer</td>
<td>Urine</td>
<td>To monitor response to treatment</td>
</tr>
<tr>
<td>Prostate-specific antigen (PSA)</td>
<td>Prostate cancer</td>
<td>Blood</td>
<td>To help in diagnosis, assess response to treatment, and look for recurrence</td>
</tr>
<tr>
<td>Thyroglobulin</td>
<td>Thyroid cancer</td>
<td>Tumor tissue</td>
<td>To evaluate response to treatment and to look for recurrence</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------</td>
<td>--------------</td>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>Urokinase plasminogen activator (uPA) and plasminogen activator inhibitor (PAI-1)</td>
<td>Breast cancer</td>
<td>Tumor tissue</td>
<td>To determine aggressiveness of cancer (and guide treatment)</td>
</tr>
<tr>
<td>70-Gene signature (Mammaprint)</td>
<td>Breast cancer</td>
<td>Tumor tissue</td>
<td>To evaluate risk of recurrence</td>
</tr>
<tr>
<td>21-Gene signature (Oncotype DX)</td>
<td>Breast cancer</td>
<td>Tumor tissue</td>
<td>To evaluate risk of recurrence</td>
</tr>
<tr>
<td>5-Protein signature (Ova1)</td>
<td>Ovarian cancer</td>
<td>Blood</td>
<td>To pre-operatively assess pelvic mass for suspected ovarian cancer</td>
</tr>
</tbody>
</table>

Table taken from [2]
4.6 MicroRNA: biogenesis, functions and targets

MicroRNAs (miRNAs) represent a large family of small non-coding RNAs of about 25 nucleotides in length that serve as effector molecules of sequence-specific gene silencing [132]. It is estimated that the number of miRNAs in the human genome range from about 450 to 1000 and that they control gene expression of about 30% of all protein-coding genes in mammals. The majority of identified miRNAs, are highly evolutionary conserved among many distantly related species, from worms to human, suggesting that miRNAs have roles in essential biological processes, including developmental timing, stem-cell differentiation, signal transduction.

In the nucleus, miRNA genes are generally transcribed by RNA polymerase II or III to form large primary miRNA transcripts (pri-miRNAs). These are further processed by Drosha, an RNase III protein, into 70-nucleotide miRNA precursors (pre-miRNA). After transport into the cytoplasm, pre-miRNAs are further processed by another RNase III enzyme, Dicer, into miRNA duplexes, typically consisting of 19-25 nucleotides in length [133, 134]. Subsequently these duplexes can be loaded into the miRNA-associated multiprotein RNA-induced silencing complex (miRISC) and the mature miRNA strand is preferentially retained. Once bound to the 3’ untranslated region (UTR) of target mRNAs, the mature miRNA induces cleavage, translational repression or deadenylation, depending on the degree of complementarity [135]. A single miRNA may bind to as many as 200 target genes encoding a broad range of proteins, such as transcription factors, receptors and transporters. In recent years, several approaches have been used to identify miRNA targets [136]. For all known miRNA targets, perfect or near-perfect complementary sites of miRNAs have been conserved. This provides a powerful strategy for the prediction of miRNA targets through computational approaches, and several laboratories have developed different computational methods to achieve this using the available genome database; hundreds of miRNA targets for given miRNAs were predicted [137].
MicroRNA biogenesis

The miRNA processing pathway, taken from Filipowicz et al. [132]: canonical maturation includes the production of the primary miRNA transcript (pri-miRNA) by RNA polymerase II or III and cleavage of the pri-miRNA by the microprocessor complex Drosha–DGCR8 (Pasha) in the nucleus. The resulting precursor hairpin, the pre-miRNA, is exported from the nucleus by Exportin-5–Ran-GTP. In the cytoplasm, the RNase Dicer in complex with the double-stranded RNA-binding protein TRBP cleaves the pre-miRNA hairpin to its mature length. The functional strand of the mature miRNA is loaded together with Argonaute (Ago2) proteins into the RNA-induced silencing complex (RISC), where it guides RISC to silence target mRNAs through mRNA cleavage, translational repression or deadenylation, whereas the passenger strand (black) is degraded. In this review we discuss the many branches, crossroads and detours in miRNA processing that lead to the conclusion that many different ways exist to generate a mature miRNA.
4.6.1 MicroRNA and cancer

Early observations suggested a potential role of miRNAs in human carcinogenesis. miRNAs discovered in *C. elegans* and in *Drosophila* were shown to control cell proliferation and apoptosis [138, 139] and their deregulation can contribute to proliferative diseases like cancer. It has also been shown that malignant tumors and tumor cell lines have widespread deregulated miRNA expression compared to normal tissues [140, 141]. One of the questions to be answered is if miRNA expression observed in cancer is cause or rather consequence of malignant transformation [142].

In the past 10 years many studies were performed in this direction, with the first direct evidence shown by Calin *et al.* They identified two miRNA, *mir*-15 and *mir*-16, transcribed from genes located in a 30-kb deletion on chromosome 13, which is the most common chromosomal abnormality in chronic lymphocytic leukemia (CLL), consequently they found the two miRNAs absent or downregulated in the majority (68%) of cases compared to normal tissue. This finding suggested that these two miRNAs were causally involved in the pathogenesis of chronic lymphocytic leukemia [143].

Other studies described specific activities of some microRNAs as oncogenes or tumor suppressors. He *et al.* showed the relationship between a miRNA cluster, *mir*-17-92 (*oncomir-1*) with the Myc oncogenic pathway [144, 145], using a mouse model of human B-cell lymphoma caused by *Myc* oncogene overexpression. They demonstrated that upregulation of the *mir*-17-92 cluster accelerated c-Myc-induced tumorigenesis in mice. The oncogenic role of the *mir*-17-92 cluster was confirmed by Matsubara *et al.* and Sylvestre *et al.* [146, 147]. On the other hand, O’Donnell *et al.* identified the same cluster of miRNAs, *mir*-17-92, to be regulated by the transcription factor c-Myc which induces expression of a growth factor E2F1. The *mir*-17-92 cluster, which is also induced by c-Myc, in contrast, inhibits E2F1 expression, conferring a potential role as a tumor suppressor [144, 145].

MicroRNAs are also involved in the regulation of p53 that regulates part of the stress responses including response to the DNA damage. miR-34, is directly activated by p53 after DNA damage. Overexpression of miR-34 induces cell cycle arrest and promotes apoptosis [148, 149].
It has been shown that expression of many miRNAs is significantly reduced in cancers compared to matched normal tissues. Furthermore, undifferentiated tumors had lower miRNA levels compared with more-differentiated tumors [141, 150], pointing to possible miRNAs functions in terminal differentiation and cell separation. Kumar et al. verified that reduction in miRNA levels promote tumorigenesis by knockdown of Drosha and Dicer in cell lines [151]. These cells with global miRNA downregulation showed enhanced cellular growth in vitro. When injected into nude mice, they generated faster growing and more invasive tumors compared to controls. To assess the effect of global miRNA dowregulation in vivo, the authors deleted Dicer in a mouse model of lung cancer. The enzyme lacking mutant mice developed an increased tumor burden, with a growth in tumor number and tumor size, as well as tumors that were less well differentiated compared to controls. These data suggest that global miRNA downregulation enhances tumorigenesis [151]. Together, these data indicate that altered expression of miRNAs has a pleiotropic effect on tumor suppressors and oncogenes and represents a crucial step in tumorigenesis.
4.7 HMG proteins

High Mobility Group (HMG) proteins, can be divided into three distinguishable families; (1) HMGA, (2) HMGB, and (3) HMGN. All three families share many biochemical features, but each has its own characteristic functional motifs and is expressed in distinctive ways in different cells and tissues [152].

HMG protein families are classified as ‘architectural transcription factors according to their ability to positively and negatively regulate gene transcription by binding to DNA or chromatin in a structure-specific manner. HMG proteins have been implicated in a diverse array of additional nuclear processes like chromatin and nucleosome remodeling events, cell cycle-related chromosomal changes, genetic recombination, DNA replication and repair, apoptosis, and molecular chaperoning [153-157].

4.7.1 HMGA family

The HMGA protein family consists of HMGA1 and HMGA2 genes that encode four proteins named HMGA1a, HMGA1b, HMGA1c (spliced isoforms of the HMGA1 gene; locus 6p21) and HMGA2. The HMGA1a, HMGA1b and HMGA2 proteins are composed of 107, 96 and 108 amino acid residues, respectively. Each protein contains three basic domains, named AT-hooks and an acidic C-terminal region, this motif bind the minor groove of a target DNA strand. The HMGA1a protein differs from HMGA1b in that it has an additional insertion of 11 amino acid residues between the first and the second AT-hook domains. The structure of HMGA2, very similar to that of HMGA1b, contained a short peptide of 12 amino acid residues between the third AT-hook and the acidic C-terminal [158].

Until 2004 there were reports of over 50 different eukaryotic and viral genes regulated by HMGA proteins. The vast majority (>35) of these are positively regulated and their inducible expression is controlled by a variety of biological and environmental stimuli. The promoter regions of many of the positively regulated genes contain multiple stretches of AT-rich sequence (AT-hook binding site). Transcriptional activation of these types of promoter often
involves the formation of an “enhanceosome”, a stereo-specific, multi-protein complex that includes HMGA proteins and other transcription factors making specific protein-DNA and protein-protein contacts in intricate, but precise, ways. In the case where HMGA proteins act as negative regulators of gene transcription they often serve as inhibitors of enhanceosome formation, usually by sterically blocking the functional binding of other crucial transcription factors to their recognition sites in gene promoters [159]. Very recently, the down-regulation of the recombination activating gene (RAG2) by HMGA1 proteins has been described [160]. One of the best-studied mechanisms of gene regulation in which HMGA proteins are involved is that of the interferon-β gene (IFN-β). IFN-β expression is due to a multifactorial complex that assembles at the nucleosome-free enhancer region of the gene, formed by the factors NF-kB, IRF, ATF2/cJun, and the HMGA1a protein. HMGA1a plays a dual function in this context: (i) it induces allosteric changes in the DNA, thus increasing the affinity of the transcription factors for their binding sites and (ii) it establishes protein-protein interactions with the same factors. This new structure, called enhanceosome, is responsible for the modification and the remodeling of a nucleosome that masks the TATA-box; consequently, transcription can start. This remodeling is triggered by the recruitment from the “enhanceosome” of GCN5/PCAF that acetylates the nucleosome and also HMGA1a at K64, the latter modification resulting in the stabilization of the enhanceosome. Later, another acetyltransferase called CBP modifies HMGA1a at K70 destabilizing the enhanceosome and, consequently, repressing transcription [161].
4.7.2 HMGA expression in normal and neoplastic tissues

The HMGA2 gene is not expressed in human tissues except for CD34 positive hematopoietic stem cells. The HMGA1 gene is expressed at low levels in human tissues: a higher expression was observed in testis, skeletal muscle and thymus. Conversely, both genes are widely expressed during embryogenesis. HMGA1 and HMGA2 over-expression was first described in rat thyroid transformed cells and in experimental thyroid tumours. Over-expression of the HMGA proteins was then found to be a common feature of experimental and human malignant neoplasias, including thyroid, prostate, uterus, breast, colorectum, ovary and pancreas carcinomas [158]. Recently, HMGA1 expression has been correlated with the histological grade of human glial tumors [162]. Moreover, the expression level of the HMGA proteins is significantly correlated with parameters of poor prognosis in patients with colorectal cancer. In all of these epithelial/endothelial cell-derived malignant tumors, the over-expressed proteins are full-length non-mutants forms. In contrast to the situation in carcinomas, benign tumors of mesenchymal origin (lipomas, leiomyomas, fibroadenomas, aggressive myxomas, pulmonary hamartomas and endometrial polyps) often contain chromosomal rearrangements that result in the creation of new hybrid genes that code for chimeric proteins in which the AT-hooks of the HMGA proteins are fused to ectopic peptidic sequences [163]. Over-expression of the HMGA proteins appears to be a necessary event in in vivo cell transformation. This was demonstrated by experiments in which HMGA expression was blocked by transfecting rat thyroid cells with antisense HMGA constructs. When these cells were infected by the myeloproliferative sarcoma virus and the Kirsten murine sarcoma virus carrying the v-mos and v-ras-Ki oncogenes, respectively, they did not acquire the typical markers of neoplastic transformation (ability to grow in soft agar and induce tumors after injection into athymic mice). Conversely, these markers were shown by the untransfected rat thyroid cells infected with the same murine retroviruses [164]. Over-expression of HMGA1 proteins is also essential in the development of cancer in humans. In fact, an adenovirus carrying the HMGA1 gene in an antisense orientation induces programmed cell death in carcinoma
cell lines derived from human thyroid, lung, colon and breast cancers. Moreover, it has been reported that the overexpression of HMGA1a or HMGA2 leads to neoplastic transformation of both Rat-1a fibroblasts and CB33 cells, whereas the decrease of HMGA1a/b expression abrogates transformation in Burkitt's lymphoma cells [165]. In chapter s 6.3 and 6.3.1 we characterized HMGA1, HMGA2, at protein level in different human neoplastic entities, correlating their overexpression with clinic-pathological features.
4.8 Cancer Stem Cell Markers

Stem cells are defined as cells that have the ability to perpetuate themselves through selfrenewal and to generate mature cells of a particular tissue through differentiation [166]. Because normal stem cells and cancer cells share the ability to self-renew, it seems reasonable to propose that newly arising cancer cells appropriate the machinery for self-renewing cell division that is normally expressed in stem cells. Evidence shows that many pathways that are classically associated with cancer may also regulate normal stem cell development [167-169]. In most tissues, stem cells are rare but it has been shown for solid cancers that these cell populations are phenotypically heterogeneous and that only a small proportion of cells are clonogenic. These observations led to the hypothesis that only a few cancer cells are actually tumorigenic and that these tumorigenic cells could be considered as cancer stem cells [166, 170-175]. The cancer stem cell hypothesis postulates that a small subpopulation of cancer cells possessing self-renewal characteristics is responsible for initiating and maintaining cancer growth. According to the CSC model the large populations found in a tumor might represent diverse stages of differentiation. The biological characteristics shared by normal stem cells (NSCs) and CSCs mainly involve self-renewal and differentiation potential, survival ability, niche-specific microenvironment requirements and specific homing to injury sites and may have important implications in terms of new approaches to cancer. The identification of new therapeutic targets based on the CSC model represents a great challenge.

In the chapter 6.4 in according to the CSC hypothesis, we evaluated a potential role of CSC proteins in tumors of the ampulla of Vater. Our findings indicate, that in ampullary carcinomas, loss of expression of EpCAM may be linked to a more aggressive tumor phenotype.
4.9 MAGE Gene Family

The melanoma antigen gene (MAGE) family is divided into two subfamilies: MAGE-I and II. MAGE-I consists of a large number of genes located on the X chromosome, including: MAGE-A located on Xq28 [176, 177], MAGE-B cluster located on Xp21 [178, 179], and MAGE-C on Xq26–27 [180]. MAGE-A, -B, and -C is characterized by a large terminal exon encoding the entire protein. Most of them are relevant cancer/testis antigens (CTA) [181] and then silent in normal adult tissues except in male germ cells [182], and highly expressed in various forms of cancer [183]. Studies of MAGE-I genes, aiming at quantifying mRNA expression in cancer (due to limited specificity of MAGE antibodies), have found the highest frequency of MAGE-I mRNA expression in different types of cancers like melanoma and lung cancer, particularly the squamous cell type. In contrast, hematopoietic malignancies, including lymphomas and leukemia as well as renal, colon, and pancreatic cancers, displayed notably low MAGE-I expression. For instance, MAGE-A3 mRNA expression has been observed in 85% of non-small-cell lung carcinoma [184], but only in 41% of multiple myeloma [185], 37% of bladder cancer [186] and 10% of the breast cancers [187]. Most investigations showed high expression of MAGEs, except MAGE-A4 that was often associated with poor outcome [188]. Higher grade and metastatic tumors have also been found to have high MAGEs expression than the primary tumors [189, 190].

The MAGE-II family, which includes the MAGE-D group, differs from the previously described members since they are almost universally expressed in all tumor free tissues and not related to cancer. They also differ by their genomic structure, the open reading frame of hMAGE-D2 being split over 11 exons. Importantly, MAGE-D1 was recently found to interact with the p75 neurotrophin receptor and to facilitate nerve growth factor-dependent apoptosis [191]. MAGE-D1 was also recently reported to interact with members of the Dlx/Msx homeodomain family and to regulate the transcriptional function of Dlx5 [192].

Encoding tumor-specific antigenic peptides directed against MAGE-I expressing cancer cells, MAGE-I protein-derived peptides are currently studied as targets for the development of cancer vaccines [193, 194] and
antitumor immunotherapy [195-197]. In addition, expression analysis in various malignancies could be of diagnostic and/or prognostic relevance. In chapter 6.5 we generated, a MAGE-A10 monoclonal antibodies (mAbs) to staining a multitumor tissue microarray. We described an overexpression of MAGE-A10 in cells from a number of malignancies, including lung, skin and urothelial tumors.
5. AIMS

My thesis focuses on the discovery and characterization of new diagnostic and prognostic markers in various cancer entities, in particular in sporadic and hereditary colorectal cancer. The work can be divided in seven major parts with the subdivision reflecting the structure of the results and the appendix.

The aims of the first part (chapter 6.1) were:

- To elucidate the frequency of contractions/insertions in the 3’UTR of the *EWSR1* gene on MSI detection in MMR deficient and proficient cancers through capillary electrophoresis analysis 319 patients.
- The role of these modifications in mRNA processing and protein expression *in vitro* (siRNA-mediated poly(A) site selection and pull-down assays) and *in vivo* (mRNA and protein expression by RT-PCR and tissue microarray analysis).

The aims of the second part (chapter 6.2) were:

- To understand the type and frequency of genetic as well as epigenetic alterations at the *SH2D4A* locus in colorectal cancer via Sanger sequencing, gene dosage analysis by multiplex ligation-dependent probe amplification (MLPA), quantitative PCR for mRNA and immunohistochemical staining for protein expression.
- To define the *SH2D4A* pathway interactions via immunoprecipitation and cell transfections.
- To test whether *SH2D4A* expression impacts on cell growth via interaction with its partners by knock-down experiments.

The aim of the third part (chapters 6.3 and 6.3.1) was:

- To investigate the prognostic role of HMGA1 and HMGA2 protein in breast and pancreatic cancers applying immunohistochemical techniques on tissue microarrays and gathering of clinic-pathological information and statistical comparison.
The aim of the fourth part (chapter 6.4) was:

• To explore the prognostic usage of cancer stem cell markers in ampulla Vater carcinomas using the tissue microarray technology followed by survival analysis.

The aim of the fifth part (chapter 6.5) was:

• To develop a specific MAGE monoclonal antibody and investigate its expression in a multitumor array by immunohistochemical techniques.

The aims of the sixth part (chapter 8.1) were:

• To assess the miRNA profile in Lynch syndrome-associated colorectal cancers (CRC) and establish their prognostic and/or therapeutic significance by deep sequencing of MMR deficient and proficient cell lines as well as cancers from Lynch syndrome and sporadic CRC patients.

• Verification and validation of selected, differentially expressed miRNAs using quantitative real-time PCR and cell transfection assays.

The aims of the seventh part (chapter 8.2) were:

• To investigate copy number variation aberrations in CRCs from six Lynch syndrome patients using a high-resolution DNA SNP array.

• To validate the findings in a cohort of 46 Lynch syndrome related and 50 sporadic CRCs.
6. RESULTS

6.1 3’UTR poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair deficient cancer.

My contribution to this work:
- RNA and DNA extraction from Lynch syndrome and sporadic colorectal cancers
- Molecular assessment of EWS16T contractions/expansions;
- Microsatellite instability analyses of Lynch syndrome and sporadic cancers;
- Quantitative real time PCR experiments;
- Immunohistochemical analysis of CRC whole tissue and microarray sections;
- Data and statistical analysis;
- Manuscript writing;
3’UTR poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair deficient cancers

1Biozentrum, University of Basel, Basel, Switzerland.
2Research Group Human Genetics, Department of Biomedicine, University of Basel, and Division of Medical Genetics, University Children's Hospital, Basel, Switzerland.
3Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki, Finland.
4Institute of Pathology, University Hospital of Basel, Basel, Switzerland.
5Institute for Surgical Research and Hospital Management and Department of Biomedicine, University of Basel, Basel, Switzerland.
6Pathologie Laenggasse Bern, Bern, Switzerland
7Institute of Molecular Cancer Research, University of Zurich, Zurich, Switzerland.

* These authors equally contributed to the work
$Corresponding authors:
Mihaela Zavolan
Biozentrum
University of Basel
Klingelbergstrasse 50 / 70
4056 Basel, Switzerland
Phone: +41 61 267 15 77
E-mail: mihaela.zavolan@unibas.ch

and
Karl Heinimann
Research Group Human Genetics
Departement of Biomedicine, University of Basel
Mattenstrasse 28
4058 Basel, Switzerland
Phone: +41 61 267 07 73
Fax: +41 61 267 07 78
E-mail: karl.heinimann@unibas.ch
Microsatellite instability (MSI), the genome-wide accumulation of DNA replication errors at repetitive nucleotide sequences, constitutes the hallmark lesion of DNA mismatch repair (MMR) deficient cancers\(^1\). Present in Lynch syndrome-related and about 10-20% of sporadic colorectal (CRC), gastric and endometrial cancers MSI testing is widely used to guide clinical management. The functional significance of MSI at non-coding repeat loci such as the 3’ untranslated region (UTR), however, remains largely elusive\(^1\)\(^-\)\(^3\). Here we describe a mononucleotide (T/U)\(_{16}\) tract, EWS16T, located in the 3’ UTR of the Ewing sarcoma break point region 1 (EWSR1) gene which discriminates MMR proficient from MMR deficient cancers with 100% sensitivity and specificity. We demonstrate \textit{in vitro} and \textit{in vivo} that contractions at this locus alter poly(A) site selection by promoting SFPQ-mediated distal poly(A) site usage in \textit{EWSR1} pre-mRNAs and result in decreased mRNA as well as protein expression. In contrast to their proficient counterparts, MMR deficient CRC display altered subcellular localization of EWS with diffuse cytoplasmic staining. EWS16T thus not only represents a novel monomorphic MSI target locus to accurately identify both, hereditary and sporadic, MMR deficient cancers but contractions therein affect multiple regulatory mechanisms implicating the RNA-/DNA-binding protein EWS in MSI-associated colorectal tumorigenesis.
Our investigation of an extra-osseous Ewing sarcoma (ES) from a MSH6 mutation carrier (c.3696dupT) previously affected by colon cancer revealed that the tumor displayed the typical hallmarks of LS, i.e. MSI at mononucleotide markers and specific loss of MSH6 expression (Supplementary Fig. S1), but none of the molecular features commonly associated with ES, i.e. chromosomal translocations involving the EWSR1 gene on 22q12 (Supplementary Fig. S2). In view of the tumor’s MSI-high phenotype we focused on a mononucleotide tract in the 3’ UTR of the EWSR1 gene consisting of 16 thymines (EWS16T; c.*318_*333). Both of the patient’s tumors, ES and colon cancer, were found to carry somatic contractions/deletions of 4 and 5 thymines, respectively, prompting us to assess the prevalence of somatic alterations at EWS16T in MMR deficient cancers in general. In contrast to Wheeler et al. who reported a deletion variant (c.*331_*333delTTT; dbSNP: rs76631619) in James D. Watson’s genome, we found the EWS16T locus to be monomorphic in over 300 constitutional DNA samples tested. As a first step, we analyzed 85 cancers (78 colorectal (CRCs) and 7 endometrial) and matching leukocyte-derived DNA samples from 78 Swiss LS patients with confirmed MMR germline mutations (58 MLH1, 20 MSH2). In addition, we investigated 14 sporadic MMR-deficient CRCs with MLH1 promoter hypermethylation as well as 85 sporadic MMR-proficient CRCs. Assessment of EWS16T tract length by capillary electrophoresis of fluorescently labeled PCR products revealed that all MMR-deficient cancers but none of the MMR-proficient ones displayed novel alleles, i.e. contractions or expansions at the EWS16T tract (Fig. 1). Subsequently, the findings were independently confirmed in a Finnish cohort of 122 patients: all 29 MMR-deficient (12 CRC, 17 gastric cancers) but none of the 93 MMR-proficient (38 CRC and 55 gastric) cancers showed EWS16T tract instability. The majority (72.7%) of somatic alterations consisted of contractions/deletions of 4 or more base pairs (Supplementary Fig. 3). Consistently, we further found that MMR-deficient cell lines (LoVo, HCT15, HCT116) carry only mutated EWS16T alleles (contractions). In contrast, MMR-proficient cell lines (HT29, SW480) as well as 12 MMR-proficient, MSI-low CRCs from Swiss patients were stable at EWS16T (Table 1). The
EWS16T tract thus represents a novel, monomorphic MSI target locus identifying MMR-deficient cancers with 100% sensitivity and specificity. Because somatic EWS16T tract alterations were exclusively present in MMR-deficient cancers, whether hereditary or sporadic, and occurred in all types of cancer investigated (colorectal, gastric and endometrial) we wondered about their possible functional role(s) in MMR-related carcinogenesis. The poly(T/U) tract deletions occur in a region which encodes the 3'UTR of EWSR1 and do not alter the coding sequence of the EWSR1 gene. Nonetheless, 3'UTRs contain sequence elements that are important for the post-transcriptional regulation of protein levels. Furthermore, it has been recently demonstrated that changes in 3'UTR length through alternative polyadenylation activates oncogenes. We thus set out to characterize the effect of EWS16T tract deletions on EWSR1 expression levels. The catalog of 3' end cleavage sites that we recently generated through 3' end sequencing in human embryonic kidney (HEK) 293 cells shows that EWSR1 undergoes alternative polyadenylation, generating two transcript forms that differ in the length of their 3'UTRs (Supplementary Fig. 4). The ESW16T tract deletions occur very close to the distal poly(A) signal and may thus result in changes in 3' end processing factor assembly, thereby altering the poly(A) site selection. To investigate this possibility, we cloned the 3'UTR of EWSR1 downstream of the Renilla luciferase-coding region in a psiCHECK-2 mutant vector in which the synthetic poly(A) site was mutated (psiCHECK-2-SPAm). This construct thus allowed only the usage of the poly(A) signal from the cloned 3'UTR of EWSR1. We then generated variant constructs containing poly(T/U) tracts of variable lengths through deletion mutation. With primers that simultaneously detect both the short and the long 3'UTR isoforms in a multiplexed semi-quantitative PCR, we found that deletions in the ESW16T tract promoted the usage of the distal poly(A) site (Figure 2a-b). We further investigated the MMR-proficient and MMR-deficient colon cancer cell lines and found that, consistent with our findings in the heterologous system, MMR deficiency is associated with higher expression of the longer EWSR1 isoform (Figure 2c). These results indicate that the EWS16T tract deletions alter poly(A) site selection. To determine the factors involved in EWSR1 poly(A) site selection, we used S1 aptamer-tagged, in vitro transcribed wildtype 3'UTR and a 3'UTR variant with 6U deletions in
the EWS16T region to pull down the proteins that associated with these RNAs (Supplementary Fig. 5). In three independent experiments we reproducibly identified a set of A/U-rich element binding proteins that associate with these constructs (Figure 3A and 3B). Interestingly, we found that NF45/90/110, hnRNPC and HuR associate with the wildtype but not with the mutant 3’UTR. NF45 and NF90 have been previously shown to be a part of a heterodimeric complex, nuclear factor of activated T-cells (NFAT), which is required for T-cell expression of interleukin 2, with NF110 being a larger isoform of NF90. NF90 has been shown to regulate mRNA stability and redistribution of nuclear mRNAs in the cytoplasm\(^7\). The EWS16T mutant preferentially associated with the SFPQ/NONO heterodimer, which is an essential pre-mRNA splicing factor that couples splicing with polyadenlyation as a component of a small nuclear ribonucleoprotein (snRNP)-free complex with SNRPA/U1A\(^8\). To determine which of the RBPs identified above influenced the poly(A) site selection, we knocked down their expression individually with siRNAs (Figure 4a) and assessed the poly(A) site usage in reporter constructs that had either the wildtype or the mutant EWS16T tract (6U deletions) cloned downstream of luciferase. None of the siRNAs influenced the poly(A) site usage in the wildtype constructs in which only the shorter isoform was expressed (Figure 4b). When the 6U deletion construct was used however, the knockdown of SFPQ and hnRNPC strongly influenced poly(A) site selection. Specifically, knockdown of SFPQ promoted proximal site usage, while the knockdown of hnRNPC led to increased expression of the longer 3’UTR isoform (Figure 4c). This result was consistent with the binding pattern of the RBPs. hnRNPC binding was most prominent immediately downstream of the proximal poly(A) site (our unpublished PAR-CLIP data on hnRNPC), while SFPQ most likely bound to the truncated U-tract immediately upstream of the distal poly(A) site, as inferred from the pull down experiments (Figure 3a). Interestingly, knockdown of SFPQ did not affect poly(A) site selection in wildtype constructs, consistent with our earlier findings that SFPQ specifically associated with the construct carrying the 6U deletion. Thus, our results indicate that hnRNPC and SFPQ have antagonistic activity on the processing of mutant \textit{EWSR1} pre-mRNA with hnRNPC promoting the generation of the shorter and SFPQ of the longer 3’UTR isoform.
To finally determine if the choice in polyadenylation site may influence EWSR1 expression, we performed luciferase assays on the constructs that carried the wildtype or various EWS16T deletion variants. The results shown in Supplementary Fig. 6 indicate a significant downregulation (up to 30%) of protein levels associated with EWS16T tract deletions. We attempted to corroborate our observations in vivo, assessing EWSR1 mRNA expression in a set of 8 LS-related MMR-deficient and 5 sporadic MMR-proficient CRCs relative to matched, tumor-free mucosa by RT-PCR. Four (50%) out of the 8 LS, but only 1 (20%) out of the 5 sporadic CRC showed significantly (>1.2 fold) increased expression of the longer 3'UTR isoform (Figure 5A). The total EWSR1 mRNA levels, however, were significantly reduced (>1.2 fold) in both LS-related and sporadic CRC tissues (Figure 5B), pointing to additional mechanisms that regulate EWSR1 expression in sporadic CRCs. To determine the consequences of altered EWSR1 mRNA expression, we performed immunohistochemical analysis (IHC) of 10 LS-related, MMR-deficient and 9 sporadic, MMR-proficient CRCs. Consistent with the data at the mRNA level, the cancers displayed on average an approx. 30% reduction in EWS expression when compared to matched, tumor-free mucosa. Unexpectedly, however, MMR-deficient and -proficient cancers significantly differed with regard to the subcellular localization of EWS (P<0.001): tumor-free colon mucosa and adenomas from LS and sporadic CRC patients as well as sporadic carcinomas showed exclusively nuclear expression (Fig. 6). In contrast, LS-related CRCs displayed diffuse cytoplasmic EWS expression (Fig. 6f). These results were subsequently confirmed by IHC analysis of a tissue microarray (TMA) containing 64 sporadic and 94 LS-related CRCs: we observed a reduction of approx. 30% in EWS expression in both groups, but only the LS-related cancers showed diffuse cytoplasmic staining for EWS (61% vs 3% of the sporadic cancers). Thus, MMR-deficient CRCs, all carrying somatic EWS16T tract alterations, display a distinct subcellular EWS distribution pattern in vivo. Further studies are needed to assess if this is directly related to EWS16T tract alterations or, rather, an indirect consequence of MSI-associated genetic instability affecting e.g. methyl-transferases like PRMT1 known to regulate the localization of EWS by methylating
Glycine/Arginine-rich motifs located in the arginine-glycine-glycine domains of EWS.

In summary, EWS16T represents a novel, monomorphic MSI target locus which identifies both, hereditary and sporadic, MMR deficient cancers with 100% sensitivity and specificity. The contractions at this locus affect multiple regulatory mechanisms including alternative polyadenylation, mRNA / protein expression and possibly subcellular localization thereby implicating the RNA-DNA-binding protein EWS, critical for the maintenance of genome integrity\textsuperscript{10}, in MSI-associated colorectal tumorigenesis.

ACKNOWLEDGEMENTS

We thank the patients and their families for participating in this study and the treating physicians for providing pertinent clinical information and tumor specimens. In particular, we are indebted to Prof. G. Spagnoli, Dept. Biomedicine, University of Basel, for sharing his expertise on cell culturing and harvesting. We also thank Michèle Attenhofer, Nemja Boesch and Sibylle Bertschin for excellent technical assistance.

The study was supported by grants from the European Molecular Biology Organization (long-term fellowship 184-2009 to S.K.), the Academy of Finland (P.P.), the Finnish Cancer Organizations (P.P.), the Sigrid Juselius Foundation (P.P.), Biocentrum Helsinki (P.P.), the European Research Council FP7-ERC-232635 (P.P.), the University of Basel (M.Z.), the Krebsliga beider Basel (K.H.), the Krebsliga Zentralschweiz (K.H.) and Oncosuisse (K.H.).
AUTHOR CONTRIBUTIONS
S.K. performed the in vitro transcription assay, protein pull down assay, poly(A) site selection assay and luciferase assays; S.P. and M.K. performed the MSI analyses on the Swiss patients and quantitative real time PCR experiments; P.P and A.G performed the MSI analysis on the Finnish patients, S.P., F.T. and L.T. analyzed CRC whole tissue and microarray sections; H.A. and G.M. performed immunohistopathological analysis and F.W. FISH analysis on the extraosseous Ewing sarcoma; V.M. cultured and provided the MMR proficient and deficient colorectal cancer cell lines; S.P., S.K., M.K., M.Z. and K.H. analyzed the data and wrote the manuscript; all authors contributed and agreed to the final version of the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.
**FIGURE LEGENDS**

**Figure 1**

**Mismatch Repair**

Proficient cancers

![Proficient cancers graph](image)

Deficient cancers

![Deficient cancers graph](image)

**Figure 1:** EWS16T tract instability in MMR deficient cancers as determined by capillary electrophoresis of fluorescently labeled PCR products. The dotted line corresponds to the wild-type allele (16 T). MMR-proficient cancers (a) colorectal, (b) gastric; MMR-deficient cancers with expansion (c) or contraction of EWS16T (d-h): (c) colorectal, MSH2 germline mutation carrier, (d) colorectal, MSH2 mutation, (e) colorectal, MLH1 promoter hypermethylation, and cancers with MLH1 mutation (f-h): (f) colorectal, (g) gastric, (h) endometrial.
**Figure 2**

**a.** Schematic diagram of the multiplexed semi-quantitative protocol for poly(A) site selection assay. A single forward primer, which was either vector specific (for analyzing poly(A) usage of reporter constructs) or EWSR1 -3'UTR specific (for analyzing endogenous EWSR1 poly(A) usage), was used along with two terminally anchored reverse primers to yield amplicons representing both long and short variants of EWSR1 3'UTR in a single PCR reaction.

**b.** Poly(A) site usage in EWSR1 3'UTR constructs with variable poly(T/U) tracts lengths cloned in pSiCHECK-2-SPAm vector along with the quantification of distal poly(A) site usage (percentages).

**c.** Poly(A) site usage in endogenous EWSR1 gene across various MMR-deficient and proficient cell lines.

**Figure 2:** Poly(A) site selection assay through multiplexed PCR. a) Schematic diagram of the multiplexed semi-quantitative protocol for poly(A) site selection assay. A single forward primer, which was either vector specific (for analyzing poly(A) usage of reporter constructs) or EWSR1 -3'UTR specific (for analyzing endogenous EWSR1 poly(A) usage), was used along with two terminally anchored reverse primers to yield amplicons representing both long and short variants of EWSR1 3'UTR in a single PCR reaction. b) Poly(A) site usage in EWSR1 3'UTR constructs with variable poly(T/U) tracts lengths cloned in pSiCHECK-2-SPAm vector along with the quantification of distal poly(A) site usage (percentages). c) Poly(A) site usage in endogenous EWSR1 gene across various MMR-deficient and proficient cell lines.
**Figure 3A**

**Figure 3**: Identification of proteins associated with the 3’UTR of EWSR1 in a pull down assay. A) Colloidal blue stained SDS PAGE gel showing proteins specifically interacting with *in vitro* transcribed, S1 aptamer-tagged EWSR1-wt or -6del 3'UTR. “No RNA” refers to the control sample where beads were incubated with cellular lysate without a prior incubation with RNA, allowing detection of any proteins interacting non-specifically with the MyOne Streptavidin beads. Bands specifically present in either wt or 6del lanes were excised and submitted to protein identification by mass spectrometry. Identified proteins have been labeled next to the corresponding bands.
Figure 3B

Figure 3: Identification of proteins associated with the 3'UTR of EWSR1 in a pull down assay. B) Western blot confirming the results of mass spectrometry. Western blot with antibody raised against HuR (a), SFPQ (b) and NF90 (c) proteins. The membrane from (b) was used to reblot with anti-NF90 in (c). Asterix in (c) depicts the background signal from SFPQ in (b).
Figure 4: Knockdown of RNA binding proteins identified from pull down assays showing that SFPQ and hnRNPC specifically modulate poly(A) site usage in EWSR1 3'UTR reporter constructs. a) Western blot showing the knock down of the RNA-binding proteins after 72 hours. Poly(A) site usage in b) EWSR1-wt-psiCHECK-2-SPAm and c) EWSR1-6del-psiCHECK-2-SPAm reporter constructs upon knock down of RNA binding proteins. d) Quantification of the relative poly(A) site usage in c).
Figure 5A

Figure 5: Quantitative Real-Time PCR showing EWSR1 expression in tumor relative to matched normal tissue. A) Difference in polyA site usage in endogenous EWSR1 mRNA (distal/proximal).
Figure 5B

Figure 5: Quantitative Real-Time PCR showing $EWSR1$ expression in tumor relative to matched normal tissue. B) difference in total $EWSR1$ mRNA levels. Samples showing more than 1.2 fold difference are indicated with an asterix.
Figure 6: Immunohistochemical staining of EWS. a) Normal colonic mucosa from sporadic CRC patient. b) Sporadic, MMR proficient colon adenoma with reduced nuclear expression. c) Sporadic, MMR proficient colon adenocarcinoma with reduced nuclear expression. d) Normal colonic mucosa from Lynch syndrome-related CRC patient. e) MMR deficient colon adenoma with reduced nuclear expression. f) MMR deficient colon adenocarcinoma with reduced nuclear and diffuse cytoplasmic expression. CRC, colorectal cancer.
Table 1: EWS16T tract instability in 355 cancer samples. Microsatellite instability was determined according to 11. MMR MSI-H: MSI-high, MSI-L: MSI-low, MSS: microsatellite stable.

<table>
<thead>
<tr>
<th>Samples analyzed</th>
<th>N</th>
<th>MSI status</th>
<th>EWS16T Status wildtype n (%)</th>
<th>EWS16T Status contraction n (%)</th>
<th>EWS16T Status expansion n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMR deficient cancers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colorectal</td>
<td>128</td>
<td>MSI-H</td>
<td>0</td>
<td>126 (98.4)</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td></td>
<td>104</td>
<td>MSI-H</td>
<td>0</td>
<td>102 (98.04)</td>
<td></td>
</tr>
<tr>
<td>MLH1 deficient</td>
<td>76</td>
<td>MSI-H</td>
<td>0</td>
<td>76 (100)</td>
<td>0</td>
</tr>
<tr>
<td>MSH2 deficient</td>
<td>28</td>
<td>MSI-H</td>
<td>0</td>
<td>26 (92.86)</td>
<td>2 (7.14)</td>
</tr>
<tr>
<td>Gastric</td>
<td>17</td>
<td>MSI-H</td>
<td>0</td>
<td>17 (100)</td>
<td>0</td>
</tr>
<tr>
<td>MLH1 deficient</td>
<td>16</td>
<td>MSI-H</td>
<td>0</td>
<td>16 (100)</td>
<td>0</td>
</tr>
<tr>
<td>MSH2 deficient</td>
<td>1</td>
<td>MSI-H</td>
<td>0</td>
<td>11 (100)</td>
<td>0</td>
</tr>
<tr>
<td>Endometrial</td>
<td>7</td>
<td>MSI-H</td>
<td>0</td>
<td>7 (100)</td>
<td>0</td>
</tr>
<tr>
<td>MLH1 deficient</td>
<td>4</td>
<td>MSI-H</td>
<td>0</td>
<td>4 (100)</td>
<td>0</td>
</tr>
<tr>
<td>MSH2 deficient</td>
<td>3</td>
<td>MSI-H</td>
<td>0</td>
<td>3 (100)</td>
<td>0</td>
</tr>
<tr>
<td>Colon adenomas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lynch syndrome-related</td>
<td>37</td>
<td>MSS</td>
<td>37 (100)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MLH1 deficient</td>
<td>10</td>
<td>MSS</td>
<td>10 (100)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MSH2 deficient</td>
<td>8</td>
<td>MSS</td>
<td>8 (100)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sporadic</td>
<td>2</td>
<td>MSS</td>
<td>2 (100)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MMR proficient cancers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colorectal</td>
<td>190</td>
<td>MSS/ MSI-L</td>
<td>190 (100)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Coloectal</td>
<td>123</td>
<td>MSS</td>
<td>123 (100)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Gastric</td>
<td>12</td>
<td>MSI-L</td>
<td>12 (100)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Gastric</td>
<td>55</td>
<td>MSS</td>
<td>55 (100)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Methods
DNA and RNA Isolation
To isolate genomic DNA and total RNA from cell lines and fresh/frozen tumor tissue samples Qiagen QIAamp DNA/RNA Mini Kit (Qiagen, Hombrechtikon, Switzerland) and for formalin fixed paraffin embedded (FFPE) tumor samples RecoverAll™ Total Nucleic Acid Isolation Kit (Ambion, Invitrogen, Carlsbad, CA, USA) were used according to the manufacturers' guidelines.

Analysis of Microsatellite Instability (MSI)
Microsatellite instability was assessed on two independent cohorts of patients encompassing 85 Lynch syndrome related cancers (78 CRCs and 7 endometrial carcinomas), 113 sporadic CRCs, including 14 cases with MLH1 promoter hypermethylation and 12 MMR proficient, MSI-low cancers, and a Finnish cohort including 8 Lynch syndrome gastric related cancers, 50 sporadic CRCs and 64 sporadic gastric cancers. Based on the recommendations of the National Cancer Institute workshop on MSI, a panel of microsatellite loci (BAT25, BAT26, D2S123, D5S346, D17S250) and two additional microsatellite markers (BAT40, MYCL1) were used to determine MSI status. The 3'UTR poly T(16) tract (EWS16T) of the EWSR1 gene (Ewing sarcoma breakpoint region 1; RefSeq: NM_005243) was amplified by PCR with the following primers: 5’- AATGTTTCATGGTTGTGATGT-3’ (forward FAM-labeled) and 5’-GAAGGATGACTCTTTATAA-3’ (reverse). PCR products were analyzed on an ABI 310 Genetic Analyzer with GeneScan Analysis V 3.1 (PE Applied Biosystems, Foster City, CA) and Genotyper 2.0 (PE Applied Biosystems, Foster City, CA) software. Fragment analysis of PCR products allowed determination of novel alleles (expansions or contractions) within the repetitive tract of a given marker. The observed expansions or contractions in EWS16T were verified using a second set of primers covering the locus (Forward primer: 5’-GCGATGTGAGTATCATTCTG-3'; Reverse primer: 5’-AGGCCGAGAAGGATGACTCT-3’) for sequencing analysis of selected samples. Sequencing reactions using the Big Dye terminator chemistry (Applied Biosystems, Foster City, CA) were performed according to the manufacturer's protocol.
Relative expression of *EWSR1* by qPCR

*EWSR1* mRNA expression on fresh frozen tissues (8 Lynch syndrome related CRCs and 5 sporadic CRCs both matched with their tumor free mucosa) was assessed using the TaqMan® Probe-Based Gene Expression Analysis (Applied Biosystems, Foster City, CA), and the *EWSR1* probe Hs01580532_g1 (Applied Biosystems, Foster City, CA). The measurements were normalized using the *HPRT1* probes Hs01003270_g1 (Applied Biosystems, Foster City, CA)\(^{13,14}\), and the fold-changes in gene expression were calculated using the standard ΔΔCt method\(^{15}\). All retrotranscriptase reactions, including no-template controls, were run on an Applied Biosystem 7900HT thermocycler. Each sample was tested in triplicate unless specified otherwise.

Immunohistochemistry

Several cohorts of patients were studied by immunohistochemical analysis of EWS. Briefly, the tissue samples of the following cohorts of patients were analysed: 37 colon adenoma (9 of which Lynch syndrome related), 19 CRCs (10 of which Lynch syndrome related) and a tissue microarray\(^{16}\). Patient data including complete follow-up were obtained by retrospective analysis of medical records, regional tumor registries and/or treating physicians. Tissue samples were obtained by surgical or endoscopic excision. Tissues sections of 4 µm sections of paraffin embedded tissue were immunostained for primary antibody against EWS (Abcam clone 84389 dilution 1:800). Staining was carried out as previously described\(^{17}\). Immunoreactivity was scored semi-quantitatively by evaluating the number of positive tumor cells over the total number of tumor cells. Nuclear immunoreactivity scores were assigned using 5% intervals and ranged from 0% to 100%. Regarding cytoplasmic expression, the staining intensity was scored as described by Allred *et al.*\(^{18}\). All samples were examined independently by three different pathologists (S.P., F.T. and L.T.), blinded to clinicopathological and molecular genetic information.
Cell lines

Five colorectal cancer cell lines from the American Type Culture Collection (ATCC, Rockville, MD) were used for this study: four three repair deficient cell lines (HCT116, LoVo, HCT15) and two mismatch repair proficient (SW480 and HT29). HCT116, HCT15, cells (ATCC, Rockville, MD) were cultured in RPMI 1640 (Invitrogen Basel, Switzerland) supplemented with 10% fetal bovine serum FBS, 1% Kanamycin sulphate, 1% GlutaMAX-I, 1% Sodium Pyruvate, 1% non Essential Amino Acids (NEAA), 1% HEPES (all from Invitrogen Basel, Switzerland) and 0.1% 2-mercapto-ethanol (Sigma-Aldrich Basel, Switzerland). HT29 cells were grown in McCoy's 5A Medium (Invitrogen Basel, Switzerland) with 10% fetal bovine serum FBS, Kanamycin sulphate and GlutaMAX-I (all from Invitrogen Basel, Switzerland). SW480 cells were cultured in L-15 Medium (Sigma-Aldrich Basel, Switzerland) with 10% FBS, 1% GlutaMAX-I and 1% Kanamycin sulphate (all from Invitrogen Basel, Switzerland). Cells were maintained at 37°C with 5% CO2.

HeLa cells at earlier passages were cultured in DMEM with GlutaMAX (Invitrogen) supplemented with 10% fetal bovine serum FBS. Cells were maintained at 37°C with 5% CO2.

Plasmids and antibodies

Synthetic poly(A) site of Renilla luciferase gene in dual luciferase psiCHECK-2 vector was mutated using primers psi-ck2polyAmutF: 5’-GCGGCGCTGGCCGCAGCTAAATATTTTATTTTTCA-3’ and psi-ck2polyAmutR: 5’-TGAAAATAAAGATATTATTTAGCTGCGCCAGCGGGCGC-3’ using QuikChange™ Site-Directed Mutagenesis Kit as per manufacturer’s instruction to generate psiCHECK-2-Synthetic polyA mutant (psiCHECK-2-SPAm). EWSR1 3’ UTR was PCR amplified from HEK293 genomic DNA using primers EWSR1_xholF: 5’-CCGACTCGAGCGGCCCATATATCTTTATTTTCA-3’ and EWSR1_notIR: 5’-ATAAGAATGCGGCCGCGAACCAGCCTTTACCTTGA-3’. The PCR amplicons were cloned into pGEM-T Easy and subsequently subcloned into psiCHECK-2-SPAm using Xhol and NotI restriction site to generate EWSR1-wt-3’UTR/psiCHECK-2-SPAm reporter. Mutant constructs with U deletions (EWSR1-2/3/4/5/6 del-3’UTR/psiCHECK-2-SPAm) and insertions (EWSR1-
2/4/ins-3'UTR/psiCHECK-2-SPAm) were introduced using standard overlap extension PCRs.

Antibodies for Western blots against NF90, SFPQ and HuR were obtained from Santacruz Biotech and Antibody against EWSR1 for Western blots and Immunoprecipitation were obtained from Abcam.

**In vitro transcription**

In-vitro transcription for pull down assay using S1 aptamer was performed using T7 RiboMAX™ Express Large Scale RNA Production System (Promega) as per manufacturer’s instructions. Region flanking the wt- and deleted poly T/U tracks in the 3'UTR of EWSR1 constructs was amplified using primers T7_EWSR1_IVT_Cf: 5’-GCTTCTAATACGACTCACTATA GGAGAAATGGGAACCCCTTGAG-3’ and EWSR1_IVT_Cr: 5’-GAACAGAGGCCGA GAAGGAT-3’ to introduce T7 promoter sequences at the 5’ end of the amplicons. S1 aptamer sequence was introduced at the 3’ end using another round of PCR using T7_EWSR1_IVT_Cf: 5’-GCTTCTAATACGACTC ACTATA GGAGAAATGGGAACCCCTTGAT GAG-3’ and EWSR1_S1apt_IVT_Cr: 5’-CATGGCCCGGCCGCGCAT ACTCTTACGC ACTTG CATGATT TGGTCGGTCCC ATGGATCGAGAACAGAGGCCGAGAAGGAT-3’.

**Protein Pull-Down Assay**

For each sample, 100ul of the MyOne Streptavidin Dynal beads (Invitrogen) was washed twice with one bead volume of solution A (DEPC-treated 0.1 M NaOH, DEPC-treated 0.05 M NaCl) and once with one bead volume of Solution B (DEPC-treated 0.1 M NaCl) and once with RNA binding buffer. The beads were resuspended in one bead volume of RNA binding buffer (100mM NaCl, 50mM Hepes 7.5, 0.5% NP-40 and 10mM MgCl2) with 100ug of in-vitro transcribed RNA with S1 aptamer sequence and incubated at 10°C for 40 minutes in a thermomixer with intermittent shaking. The beads were washed twice with one volume of RNA wash buffer prior to incubation with the lysate. HEK293 cell pellet from 15cm² dish was lysed in 3ml native lysis buffer (25mM Hepes-KOH pH 7.5, 100mM KCl, 0.5% NP-40, 5mM MgCl2, 0.5mM DTT, protease inhibitor cocktail, 1mM NaF, 1 mM Na4VO4 and 300U of RNasin) for
15 minutes on ice. The lysate was subsequently gently sonicated and centrifuged to remove any cell debris. 200ug of *E. coli* tRNA was additional added to prevent non-specific binding of proteins to the beads. 1 ml of the lysate was added to the beads coupled to S1 aptamer RNA and also to the beads alone for no RNA control. The mixture was incubated at 4°C on rotation wheel. After 1 hour of incubation, the beads were washed thrice with with Native lysis buffer. The bound proteins were eluted with 100ul Native lysis buffer supplemented with 25mM Biotin for 30-45 minutes at 10°C on a thermomixer with intermittent shaking. 900ul of 100% ethanol was added to the eluate and incubated at -80°C for 2 hours followed by centrifugation to precipitate the eluted proteins. The pellet was air dried and dissolved in 35ul of (SDS loading dye). Prior to loading on the Nuvex gradient gels, the samples were heated at 90°C for 5 minutes. After SDS-PAGE electrophoresis, the gel was stained with colloidal blue and bands of interest were excised and sent for mass spectrophotometry.

**Poly(A) site selection assay**

HeLa cells were transfected with psiCHECK-2-SPAm constructs for 24 hours. Total RNA was isolated from the HeLa cells using TriReagent (Sigma) followed by DNAse I (Promega) treatment according to the manufacturer’s protocol. Reverse transcription was done with oligo d(T)\(_{18}\) primers for 1 hour. For poly(A) site selection assay, multiplexed polymerase chain (PCR) reaction was set up using a single forward primer specific to the psiCHECK-2 vector (psiCHECK-2-SeqFor: 5’-ATGAAATGGGTAAGTACA-3’) and two 4 nucleotide terminally anchored oligo d(T)\(_{16}\) reverse primers specific to the two isoform variants of EWSR1 (EWSpolyAproxR: 5’-TTTTTTTTTTTTTTTTTTTACCA-3’ and EWSpolyAdistR: 5’-TTTTTTTTTTTTTTTTTTTTGACT-3’ respectively) to detect only the isoforms generated specifically from the psiCHECK-2-SPAm constructs. PCR was run for 28 cycles and the products were separated on a 2% Agarose gel. The bands were quantified using the ImageJ software (http://rsbweb.nih.gov/ij/).
Luciferase assays
HeLa cells were seeded in a 48 well plate one day prior to transfection. 0.2ug of plasmids (psiCHECK-2-SPAm constructs) were transfected with Lipofectamine 2000 (Invitrogen) for 24 hours. Luciferase assays were done on the transfected cells using The Dual-Luciferase® Reporter Assay System (Promega). Both transfections and luciferase assays were done according to the manufacturer’s protocol.

siRNA transfections
Control-siRNAs and siRNAs against hnRNPC, HuR, NF90 and SFPQ and were obtained from Santa Cruz Biotechnology. HeLa cells were reverse transfected with siRNA oligos using RNAiMAX (Invitrogen). After 48 hours the cells were transfected with EWSR1-wt-3'UTR- and EWSR1-6del-3'UTR-psiCHECK-2-SPAm-constructs for another 24 hours. The cells were subsequently harvested and split into two aliquots. One aliquot was used to assess the knockdown efficiency of siRNA using Western blot, while the other was used for RNA isolation and subsequent poly(A) site selection assay.

Statistical Analyses
For statistical analysis, the chi-square test ($\chi^2$ test) and Fisher's exact test for nonparametric variables and Student’s t-test for parametric variables were used, with all probabilities reported as 2-tailed, considering a $P<0.05$ to be statistically significant. Calculations were performed using the software program SPSS 17.0 (IBM Corporation, Somers, NY 10589).

Ethical approval
The study is part of the so-called “Basler Studie über familiäre Tumorkrankheiten”, Ref.Nr.EK: 258/05 and has been approved by the “Ethikkommission beider Basel”. Furthermore, written informed consent was obtained from all Lynch syndrome patients as well as from the sporadic patients.
References


12. Boland, C.R., *et al.* A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition:


Supplementary Figure 1: Immunohistochemical staining of DNA mismatch repair proteins MLH1, MSH2, MSH3, MSH6 and PMS2 in an extra-osseous Ewing sarcoma of a MSH6 mutation carrier demonstrating selective loss of MSH6 expression.
Supplementary Figure 2

Figure S2

Supplementary Figure 2: Representative extract from fluorescence in situ hybridization (FISH) analysis of the MHS6–related extra-osseous Ewing sarcoma specimen using an EWSR1-specific break-apart probe (Poseidon™ probe KI-10708 EWS Break). Screening of 200 nuclei gave no evidence for the presence of a rearrangement involving the 22q12 locus (i.e. no evidence for any split red-green signals).
Supplementary Figure 3

Figure S3

Supplementary Figure 3: Frequency distribution of EWS16T repeat alterations in 128 MMR deficient cancers (104 colorectal, 17 gastric and 7 endometrial cancers). Numbers on top of the bars represent percentages. The dotted line corresponds to the wild-type allele (T)16.
Supplementary Figure 4: The 3'UTR of the EWSR1 gene harbors two poly(A) signals resulting in the generation of EWSR1 isoforms with short and long UTRs. 

a) Reads (per million) obtained from poly(A) seq\(^7\) showing the location of the two polyA sites in HEK293 cells, with the proximal poly(A) site being predominantly used over the distal.

b) Schematic diagram showing the location of the EWS16T tract with respect to the distal poly(A) signal (AAUAAA), and the binding specificity of important 3' end processing factors.
Supplementary Figure 5: Schematic diagram depicting essential steps of the protein pull down assay using \textit{in vitro} transcribed S1 aptamer-tagged RNA. For a more detailed protocol please refer to the Materials and Methods section of the manuscript.
Supplementary Figure 6: Luciferase assay showing the changes in Renilla luciferase protein expression from psiCHECK-2-SPAm constructs harboring different EWS16T tract deletion variants in its multiple cloning site. The signals are normalized against internal Firefly luciferase.

6.2 The 8p21.3 encoded SHOCA-2 acts as a tumor suppressor in colorectal cancer via repression of STAT3 activation.

My contribution to this work:
- DNA extraction from sporadic colorectal cancers;
- SH2D4A mutation analysis by Sanger sequencing of CRC samples;
The 8p21.3 encoded SHOCA-2 acts as a tumor suppressor in colorectal cancer via repression of STAT3 activation

Sébastien Loeffler\textsuperscript{1,2}, Michal Kovac\textsuperscript{1,2}, Stefan Weis\textsuperscript{1}, Salvatore Piscuoglio\textsuperscript{1,2}, Saulius Zuklys\textsuperscript{1,2}, Marcel Keller\textsuperscript{1,2}, Katrin Hafen\textsuperscript{1,2}, Daniel Hess\textsuperscript{3}, Kaspar
Truninger¹, Inti Zlobec⁴, Luigi Terracciano⁴, Karl Heinimann¹,², Primo Schär¹ and Georg A. Holländer¹,²,⁵

Affiliations
1: Department of Biomedicine - Mattenstrasse, University of Basel, 4058 Basel, Switzerland
2: Basel University Children's Hospital, 4058 Basel, Switzerland
3: Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland
4: Institute of Pathology, University of Basel, 4031 Basel, Switzerland
5: The Department of Paediatrics and The Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom
6: Current address: Actelion Pharmaceuticals Ltd, Allschwil, Switzerland

Corresponding author: Georg A. Holländer (georg-a.hollaender@unibas.ch)
Tel: +41 61 695 30 69 / +44 1865234 225

Running title: SHOCA-2 is a tumor suppressor in colorectal cancer
Summary

Chromosomal deletions at 8p have been associated with colorectal cancer (CRC), but the tumor suppressor function affected has not been described. Here, we identify SH2D4A as a suppressor of CRC mapping to 8p21.3. SH2D4A encodes SHOCA-2, a SH2 domain-containing adapter protein that, upon EGF signaling, interacts with the serine/threonine phosphatase PP1b and consequently inhibits STAT3 activity. Conversely, knock-down of SHOCA-2 stimulates c-Myc, Cyclin D1 and Jun B expression and accelerates CRC growth through unopposed STAT3 activity. Loss of SHOCA-2 expression in CRC correlates with advanced disease stages and poor prognosis and is caused by chromosomal rearrangements, point mutations and epigenetic changes. Thus, SHOCA-2 functions as an inhibitor of the EGFR pathway and its absence impacts on CRC development and progression.

Significance

Colorectal cancer (CRC) yearly claims more than 655,000 lives worldwide and represents the third most frequent cause of death in the Western world. Though a 8p21-23 chromosomal deletion has been linked to CRC development and progression, the molecular mechanisms related to this genomic region have not been recognized to date. Encoded on 8p21.3, the SH2 domain-containing SHOCA-2 protein represents a hitherto unrecognized CRC tumor suppressor that limits cell proliferation via a negative feedback loop controlling EGFR signaling. Consequently, a loss of SHOCA-2 expression results in the oncogenic activation of STAT3, a hallmark of many CRC. Since drugs inhibiting EGFR and STAT3 activity offer only limited efficiency, enhancing SHOCA-2 function constitutes a novel insight for the treatment of CRC.
Introduction

Oncogene activation and loss of tumor suppressor activity are responsible for cancer initiation and progression (Vogelstein and Kinzler, 2004). Overexpression of the epithelial cell growth factor receptor (EGFR) is frequently observed in colorectal cancer (CRC), and correlates with a poor clinical outcome (Saif, 2010). EGFR is receptor tyrosine-kinase that relays its activation through the signal transducer and activator of transcription 3 (STAT3) to the nucleus controlling the transcription of genes involved in cell proliferation and survival (Aggarwal et al., 2009). Constitutive activation of the STAT3 signaling pathway as a consequence of genetic mutation is oncogenic, and a hallmark of many CRC (Bromberg et al., 1999; Klampfer, 2008).

Genetic and epigenetic alterations in the expression of tumor suppressor genes are typical features of neoplastic transformation and contribute to both initiation and progression of tumor formation (Weinberg, 1991; Vogelstein and Kinzler, 2004). The loss of heterozygosity (LOH) across variable parts of chromosome 8p has been identified as a frequent abnormality associated with CRC (Emi et al., 1992) and linked to DNA breakage at fragile sites located at 8p12 and 8p22 (Birnbaum et al., 2003). Although predicted from chromosome transfer studies (Tanaka et al., 1996), specific tumor suppressor genes have to date not been identified in this critical chromosomal region. It is thus unknown how genetic alterations in 8p contribute to CRC development and progression.

Here we report on the identification of a new tumor suppressor, termed SHOCA-2. Encoded on 8p21.3, SHOCA-2 interacts with the serine/threonine phosphatase PP1b and consequently inhibits STAT3 activity to control EGFR signaling. Moreover, we characterize the genetic and epigenetic mechanisms that result in a silencing of SHOCA-2 in CRC.
Results

Identification of SHOCA-2 and its loss of expression in colorectal cancer

cDNA Representational Difference Analysis (cDNA-RDA) comparing epithelial cells at two sequential stages of embryonic maturation identified transcripts of a previously uncharacterized gene, designated SH2D4B (NCBI gene ID 387694; S.Z. and G.H., unpublished data). Homology searches identified SH2D4A as a paralogue (NCBI gene ID 63898; Figure 1A). Both genes are evolutionary conserved (Figure S1A) and ubiquitously expressed, with SH2D4A being significantly more abundant than SH2D4B (Figure S1B). The proteins encoded by SH2D4A and SH2D4B display 65% amino acid sequence identity at their C- and N-termini and share several structural motifs including three coiled-coil domains in the middle of the molecule and a Src homology (SH2) domain at the C-terminal end (Figure 1A; Figure S1C). It is for these motifs and the putative function as an adaptor that we named this family of proteins SHOCA, an acronym for SH2-domain containing adaptor protein.

SH2D4A maps to human chromosome 8p21.3 (Figure 1A) and encodes SHOCA-2. Because the chromosomal region surrounding SH2D4A (8p21-23) is frequently lost in CRC and other epithelial tumors (Emi et al., 1992), we determined by immunohistochemistry the expression of SHOCA-2 in 400 consecutive, sporadic CRCs (Figure S1D). The significant differences in detectable SHOCA-2 protein correlated with the separate clinical tumor stages (I-IV) (Figures 1B and 1C; Table S1A and S1B); SHOCA-2 expression was reduced or lost of with advanced CRC pathology. Patients with tumors displaying either low or absent SHOCA-2 expression had a poorer outcome (log-rank test \( p=0.0118 \); Figure 1D; Table S1C) though SHOCA-2 expression, when stratified by disease stage, did not serve on its own as a strong predictor of survival. An identical conclusion was drawn from a Cox regression analysis \( p=0.0134 \) with HR (95%CI)=0.56 (0.35-0.89) and a Wilcoxon test \( p=0.0133 \) (Table S1C). Gene expression data available in public databases for carcinomas of different origin including bladder, breast, head and neck, esophagus, liver and ovary suggested similar correlations between disease
stage and SHOCA-2 expression (Figure S1E). Finally, further analysis of SHOCA-1 expression (mRNA levels) in CRC samples with absent or reduced SHOCA-2 expression revealed a concurrent loss of SHOCA-1 in approximately half of the samples whereas SHOCA-1 loss was neither observed in healthy mucosa nor in CRC tissues proficient for SHOCA-2 (Figure S1F).

**Genetic and epigenetic alterations of the SH2D4A locus in CRC**

To determine the molecular basis for decreased SHOCA-2 expression, we chose from an unselected cohort of 70 CRC patients 27 subjects with SH2D4A alleles that could be distinguished by microsatellite markers and SNPs (Figure S1D). In 17 of these (63%), the primary tumor had lost or diminished SHOCA-2 expression whereas only 3 samples (11%) displayed an EGFR gene amplification. A loss of the SH2D4A gene and a simultaneous amplification of EGFR were noted in 2 out of these 3 individuals. This result is in accordance with the observation that EGFR over-expression in CRC is rarely caused by EGFR gene amplification (Saif, 2010). Using 3 flanking microsatellite markers and 6 gene-specific SNPs (Figure S2A), we detected LOH at SH2D4A in 7 of the 17 tumors (41%). Gene dosage quantification of the short arm of chromosome 8 indicated mono- and biallelic deletions in 6 and 1 of these LOH tumors, respectively (Figure 2A). When normal mucosa was examined, four of the six patients with a monoallelic SH2D4A loss in tumor tissue were heterozygous for the intronic SNP rs17128221 (c.342-5T>C). This SNP is located in a mRNA splice acceptor site and causes non-canonical exon 4 skipping, which in turn causes a translational frame shift and premature termination in exon 5 (Figure 2B). Consistent with a pathological effect, the cancer tissue of patients heterozygous for the C allele uniformly demonstrated a selective loss of the T allele (Chi square p=0.006) whereas homozygosity for the C allele was never observed in mucosa tissue samples of 83 healthy Caucasian individuals (Figure 2B).

Missense mutations and deletions in the SH2 domain of SHOCA-2 were observed in three of the 10 SH2D4A biallelic CRC samples characterized by a low or absent SHOCA-2 expression. These alterations either produced single
amino acid changes (p.Arg324Trp and p.Ser430Phe) with predicted structural effects (Polyphen-2 score of >0.95, SIFT score<0.02), or caused a translational frame shift (p.Ile378fsX15) giving rise to a loss of amino acids 377 to 454 and reduced protein expression (Figure 2C).

To determine causes for low or absent SHOCA-2 expression other than LOH and gene mutations, we next analyzed primary CRC tissue and surrounding healthy mucosa for DNA cytosine methylation in the 5' untranslated region (5'UTR) of SH2D4A (Figure 2D). Pyrosequencing of bisulfite-converted genomic DNA (Bettstetter et al., 2007) revealed two specific CpG dinucleotides showing increased methylation in cancer tissue when compared to matched normal mucosa. One of these CpGs is within a canonical Sp1 transcription factor recognition motif and its methylation significantly reduced the Sp1 binding affinity (Figure S2B). Epigenetic changes at the DNA level may therefore represent another cause for reduced SH2D4A gene expression in CRC. Taken together, in 15 of 17 CRCs investigated in molecular detail (Figure 2E), distinct genetic and epigenetic alterations explaining the loss or reduction in SHOCA-2 expression could be identified: chromosomal deletions at 8p21.3, genetic mutations in the SH2D4A gene, a mRNA splicing defect associated with a specific SNP, and site-specific changes in promoter DNA methylation.

**SHOCA-2 negatively regulates EGF-induced STAT3 phosphorylation and transcriptional activity**

A proteomic analysis of phospho-tyrosine signaling in non-small-cell lung cancer cell lines driven by EGFR-activating mutations identified a target peptide with sequence homology to SHOCA-2 (Guo et al., 2008). We confirmed that EGFR activation triggered tyrosine phosphorylation of SHOCA-2 in HeLa cells (Figure 3A). To investigating a potential involvement of SHOCA-2 in the activation of the downstream EGFR effectors STAT3, ERK and PI3K-Akt, we stimulated HeLa cells with EGF. The ectopic expression of SHOCA-2 in these cells specifically decreased the STAT3 phosphorylation at Tyr705 (Figure 3B). The phosphorylation of this residue constitutes an important post-translational modification of STAT3 as it triggers its
dimerization, nuclear translocation and DNA binding (Heinrich et al., 2003). By contrast, the relative level of Akt phosphorylation was increased whereas the MAP kinase pathway remained unaffected in EGF simulated HeLa cells (Figure S3A). Because SHOCA-2 had previously been suggested to be associated with the ER alpha/PKC signaling pathway (Li et al., 2009), we were unable to observe in ERa-negative HeLa cells (Arao et al., 2011) overexpressing SHOCA-2 a change in PKC phosphorylation further confirming the regulation of PKC by SHOCA-2 to be dependent on ERa (Figure S3B).

To analyze EGF signaling in cells with reduced SHOCA-2 levels, we generated short hairpin RNA (shRNA) expressing HeLa cell clones in which SHOCA-2 protein was knocked down by 90% (SHOCA-2 KD; Figure 3C). When stimulated with EGF, the extent of STAT3 phosphorylation at tyrosine 705 (Tyr$^{705}$) and serine 727 (Ser$^{727}$) was inversely correlated with the SHOCA-2 protein level (Figure 3C). In keeping with a significant increase in phosphorylated STAT3, the transcription of the c-Myc, Jun B and Cyclin D1 was enhanced (see below). In contrast, the phosphorylation of AKT was decreased while that of ERK 1/2 remained unchanged in SHOCA-2 KD cells (Figure S3C). These results further corroborated the specific involvement of SHOCA-2 in STAT3-mediated signaling following EGF stimulation.

Using a STAT3-driven luciferase reporter assay (Kreis et al., 2007), we next examined in HeLa cells whether STAT3 transcriptional activity was altered in the presence of decreased or increased SHOCA-2 protein levels. As demonstrated in Figure 3D, EGF treatment of HeLa cells with a significant loss of SHOCA-2 expression (SHOCA-2 KD cells) resulted in a robust transcriptional activity. In contrast, the transcription of the luciferase reporter was significantly reduced in HeLa cells overexpressing SHOCA-2 and exposed to EGF (Figure 3D; Figure S3D). To investigate the nature of SHOCA-2’s influence on STAT3, we also tested porcine aortic endothelial cells (PAE) that lack endogenous EGFR expression and mouse embryonic fibroblasts (MEF) rendered deficient for STAT3 expression (Huang et al., 2007; Costa-Pereira et al., 2002). In these cells, SHOCA-2 bound to STAT3
and EGFR and this interaction was always independent of the missing binding partner normally present in wild type cells (Figure 3E, Figure S3E). Moreover, the binding of SHOCA-2 to EGFR and STAT3 was unrelated to its SH2 domain (Figure S3F) and isoform-specific because SHOCA-1 did not physically bind to STAT3 (Figure S3G). Taken together, these results suggest that SHOCA-2 inhibits EGF-induced STAT3 activation through its physical association with both EGFR and STAT3 independently of its SH2 domain.

**SHOCA paralogues bind PP1b**

Using tandem affinity purification (TAP) and reciprocal immunoprecipitation, we and others (Ewing et al., 2007) identified the serine/threonine Protein Phosphatase-1 beta (PP1b) as an interaction partner of SHOCA-1 and SHOCA-2 (Figure 4A; Figures S4A-C; Table S2). The eukaryotic PP1 protein family is composed of the 3 isoforms a, b and g, whose functionality relies on the association with different regulatory proteins (Ceulemans and Bollen, 2004). Both SHOCA paralogues bound PP1b but neither PP1a or PP1g, demonstrating an isoform-specific interaction (Figure 4B). Confocal immunofluorescence microscopy and immunoprecipitation of PP1b from cytoplasmic and nuclear protein fractions detected the SHOCA-2-PP1b interaction in both subcellular compartments further confirming their physical association (Figure S4D).

The \([R^2K]-X(0,1)-[V^2I]-[P]-[FW]\) sequence has previously been identified as a PP1-interacting motif (Hendrickx et al., 2009). We therefore tested whether the KXILF and KX\([V^2I][Q^2H]\)W sequences located in the N-terminus of SHOCA-1 and -2 may mediate such binding (Figure 4C left panel; Figure S4E). Site-directed mutagenesis was used to alter in the SHOCA isoforms the first motif from KXILF to AXALA (designated PP1\(_{Mut1}\)) and the second sequence from KX\([V^2I][Q^2H]\)W to AXA\([Q^2H]\)A (PP1\(_{Mut2}\); Figure 4C). Overexpression of SHOCA mutant proteins in HEK293 cells demonstrated that only the PP1\(_{Mut2}\) proteins failed to interact with PP1b identifying the KX\([V^2I][Q^2H]\)W sequence as the relevant docking site (Figure 4C, middle and right panels).
STAT3 inhibition by SHOCA-2 requires EGF-induced phosphorylation of Tyr131 and the recruitment of PP1b

We next investigated whether EGFR and/or STAT3 could pre-associate with SHOCA-2 independently of any EGFR stimulation, or, alternatively, whether these interactions required an activation-induced phosphorylation of the individual binding partners. Immunoprecipitation experiments showed that SHOCA-2-flag and endogenous EGFR, STAT3 and PP1b co-precipitated in serum starved, unstimulated HeLa cells independent of prior EGF stimulation (Figure 4D). Following EGF activation, the association of SHOCA-2, STAT3 and PP1b significantly increased in HeLa cells (Figure 4D). Although STAT3 can be indirectly activated through c-SRC (Quesnelle et al., 2007), c-SRC was not associated with the SHOCA-2, STAT3 and PP1b complex demonstrating a SHOCA-2-related, direct inhibition of STAT3 that was independent of a c-SRC engagement (Figure S4F).

Following EGF stimulation, STAT3 is recruited to the phospho-sites Tyr1068 and Tyr1086 within EGFR (Shao et al., 2003). However, in Figure 4D, we demonstrated that EGFR and STAT3 associated prior to EGFR-mediated stimulation. To further characterize this interaction, unstimulated HeLa cells were transfected with mutant forms of either EGFR or STAT3. The EGFR mutant had its Tyr1068 and Tyr1086 replaced by phenylalanine (EGFR Y1068/1086F) and hence lost a motif known for classical STAT3 recruitment (Shao et al., 2003), whereas Tyr705 was replaced in the STAT3 mutant to a phenylalanine thus eliminating several essential functions including STAT3 dimerization, translocation to the nucleus and transcriptional activity (Heinrich et al., 2003). Neither the association of mutant EGFR with wild type STAT3 nor the interaction of altered STAT3 with EGFR were affected by these changes (Figure S4G), thus demonstrating that EGFR’s Tyr1068/Tyr1086 and STAT3’s Tyr705 were not critically important for the interaction with each other (Figure S4G). The latter finding is in accordance with a pTyr 705 independent recruitment of STAT3 to the IL-22 receptor, an interaction that however requires the coiled-coil domain of STAT3 (Dumoutier et al., 2009).
Because EGF-stimulation resulted in Tyr^{131} phosphorylation of SHOCA-2 (Figure 3A and (Guo et al., 2008)), we next examined whether this modification is essential for the formation of the SHOCA-2/EGFR/STAT3/PP1b complex and its function. Though changing the tyrosine residue to alanine (Y131A) did not disturb SHOCA-2’s ability to complex with EGFR, STAT3 and PP1b (Figure S4H), the point mutation resulted in a constitutive phosphorylation of other SHOCA-2 tyrosine residues and an insensitivity to EGFR signal-mediated inhibition of STAT3 tyrosine phosphorylation (Figure S4I). The regulation of STAT3 activity by EGFR is therefore critically dependent on the phosphorylation of SHOCA-2’s Tyr 131.

We next characterized in further detail the conditions under which SHOCA-2 can bind to its partners. Since both wild-type and PP1b interaction-deficient SHOCA-2 variants formed a complex with EGFR/STAT3 (Figure S4H), PP1b binding does not appear to be a prerequisite for the association of SHOCA-2 with STAT3. However, the inactivation of STAT3 was dependent on SHOCA-2’s ability to associate with PP1b as the overexpression of SHOCA-2 PP1_{Mut2} failed to alter STAT3 Tyr^{705} phosphorylation following EGF stimulation (Figure 4E). Consequently, STAT3-dependent transcription in a reporter assay was not diminished in HeLa cells overexpressing SHOCA-2 PP1_{Mut2} that had been treated with EGF (Figure 4F). SHOCA-2 thus binds to the EGFR/STAT3 complex independently of PP1b, but requires the association with the phosphatase to modulate STAT3 activity.

**SHOCA-2 requires PP1b to inhibit STAT3-dependent tumor cell growth**

To address the contribution of SHOCA-2 to cellular growth control, we assessed the expression of STAT3-controlled cell proliferation genes (Aggarwal et al., 2009; Trenerry et al., 2007) in HeLa cells expressing reduced levels of SHOCA-2 (SHOCA-2 KD). The increase of c-Myc, Cyclin D1 and Jun B transcripts in these cells upon EGF stimulation inversely correlated with a robust reduction in SHOCA-2 expression (Figure 5A, left panel). Consequently, the fraction of cells in G2/M of the cell cycle was significantly higher in both unstimulated (8.8% vs. 5% G2/M cells [p<0.0002]) and EGF activated SHOCA-2 KD HeLa cells (21.2% vs. 9.4% G2/M cells, [p<0.004])
when compared to wild type controls (Figure 5A, right panel). A significantly increased cell proliferation (Ki-67 positivity) and higher levels of c-Myc, Jun B, and Cyclin D1 transcripts were also detected in tissue sections of clinically advanced CRC further confirming an inverse correlation between SHOCA-2 expression and STAT3 mediated cellular responses (Figure 5B; Figure S5A). A decrease in SHOCA-2 expression therefore promotes the transcription of cell proliferation factors that in term drive cell cycle progression.

To determine the molecular mechanism by which SHOCA-2 suppresses cell proliferation, we next transfected the colorectal cancer cell line SW480 to over-express STAT3 alone or in combination with either wild-type or mutant SHOCA-2. In a colony formation assay, overexpression of only STAT3 significantly increased the number of colony forming units (Figure 5C), a response that correlated with a higher degree of STAT3 phosphorylation (Figure 5D). Overexpression of STAT3 together with wild-type SHOCA-2 significantly reduced the colony forming potential of SW480 cells as well as STAT3 phosphorylation. This growth suppressive effect was less pronounced when SHOCA-2 mutants deficient in either PP1b interaction (SHOCA-2 PP1Mut2) or Tyr131 (Y131A) phosphorylation were co-overexpressed instead (Figure 5C). Comparable results using the overexpression of wild type and mutant SHOCA-2 were obtained in the adenocarcinoma cell line H1975 which expresses an EGFR mutation constitutively activating STAT3 (Figure S5B) (Lu et al., 2007). Consistent with a role of SHOCA-2 in suppressing STAT3-mediated cell growth, SW480 cells in which SHOCA-2 expression was knocked down by shRNA diminished their proliferation rate upon exposure to pharmacological inhibition of STAT3 by S31-201, a chemical probe blocking STAT3-STAT3 complex formation and STAT3-DNA binding (Figure S5C). Finally, overexpression of SHOCA-2 in a CRC cell line (SW620) with constitutive STAT3 phosphorylation (Maa et al., 2007) and spontaneously low endogenous SHOCA-2 expression (Figure S5D) suppressed anchorage-independent cell growth (Figure 5E). Taken together, these results demonstrate under different experimental conditions that SHOCA-2 acts as a suppressor of STAT3-driven cancer cell proliferation.
Downregulation of SHOCA-2 alters the cell phenotype and promotes *in vivo* tumor growth

The cobble-stone shape of SW480 cells, which usually express high endogenous SHOCA-2 levels, changed to a spindle-like morphology following the knock-down of SHOCA-2 transcripts (Figure 6A, Figures S5D and S6A). This change in cell shape was paralleled by features of epithelial-mesenchymal transition (EMT) (Christofori, 2006) such as a reduction in E-cadherin and an increase in N-cadherin, vimentin, SNAIL and ZEB1 expression (Figure 6B). These alterations were, however blocked by the inhibition of EGFR using Tyrphostin (AG1478; Figure S6B). Moreover, SHOCA-2 deficient SW480 cells proliferated at a greater rate and displayed a higher sensitivity to EGFR inhibition when compared to mock transfected controls (Figure 6C). Re-establishing mouse Shoca-2 expression in human SHOCA-2 deficient SW480 cells however decreased their proliferation rate (Figure 6D), thus demonstrating again that SHOCA-2 controls STAT3-mediated tumor growth.

To establish an in vivo role for SHOCA-2 as a tumor suppressor, SW480 cells that had their SHOCA-2 expression either knocked-down or left unchanged were transplanted into nude mice. Grafts with cells where SHOCA-2 was reduced displayed a significant increase both in tumor incidence and size when compared to control transplants (Figure 6E). Moreover, significant levels of phosphorylated, nuclear STAT3 were detected in tumors that had emerged from grafted cells lacking regular SHOCA-2 expression (Figure 6F, upper left panel). Consistent with this experimental finding, increased STAT3 phosphorylation at residues Tyr<sup>705</sup> and Ser<sup>727</sup> (Figure 6F, upper right and lower panels) were also detected in the majority of CRC biopsies with low SHOCA-2 expression (H-score <100). Taken together, experimental *in vivo* tumor models using CRC cell lines and the detailed analysis of biopsies taken from CRC patients identified the inhibitory role of SHOCA-2 on STAT3 activation.
Discussion

Genomic deletions within the 8p21-23 region constitute a characteristic cytogenetic feature of CRC (Emi et al., 1992) and have been linked to carcinogenesis (Macartney-Coxson et al., 2008). How deletions in this chromosomal region contribute to CRC development has remained unclear since tumor suppressor genes could so far not be allocated to 8p21-23 (Macartney-Coxson et al., 2008; Emi et al., 1992). With the assignment of the SHOCA-2 encoding SH2D4A gene to 8p21.3, we have linked genomic instability at this locus with the loss of a specific tumor suppressor activity in CRC. We provide experimental evidence and clinical data that a loss of SHOCA-2 expression results in unopposed STAT3 activation following EGFR stimulation and marks in patients with colorectal cancer accelerated tumor growth as well as poor prognosis. Our findings of a role for SHOCA-2 in the control of tumor growth also extend to malignancies where the mechanism of tumor development and progression are still only poorly understood (Roessler et al., 2011).

EGFR is an upstream activator of multiple pathways involved in cell proliferation, apoptosis and carcinogenesis. Overexpression of EGFR and constitutive activation of its major downstream effector, STAT3, have been linked to cancer progression, a higher risk for metastasis and reduced survival (Quesnelle et al., 2007; Klampfer, 2008; Saif, 2010). Since STAT3 regulates cell growth and tissue homeostasis (Aggarwal et al., 2009), its activation must be under stringent control. We found that SHOCA-2 cooperates with PP1b in modulating the EGFR-induced STAT3 activation, thereby limiting cell proliferation. In view of the specific genetic and epigenetic alterations of SH2D4A, that particularly emerge in late stage CRCs, and given the correlation between poor clinical outcome and the loss of SHOCA-2 expression, we conclude that SHOCA-2 is a novel tumor suppressor modulating the activation of the EGFR signaling pathway (Walther et al., 2009). However, in the cohort studied here, the SHOCA-2 status was not a reliable predictor of survival in patients stratified by disease stage. This result is not surprising as it is consistent with the predictive potential of other tumor
suppressors. For example, p53, APC, and DCC also fail to serve as prognostic markers though their loss of expression in colorectal cancer is correlated with increased tumor aggressiveness (Roth, 1999; Walther et al., 2009).

Based on the evidence that SHOCA-2 binds to both EGFR and phosphatase PP1b and is itself phosphorylated upon EGF signaling, we propose that SHOCA-2 controls STAT3 activity through the catalytic activity of PP1b. SHOCA-2 and PP1b form a complex with STAT3 in that SHOCA-2 acts as an adaptor to bring the PP1b phosphatase into physical proximity of STAT3, which in turn effects the dephosphorylation of activated STAT3. Indeed, phosphorylation/dephosphorylation of Ser^{727} plays an important role in modulating STAT3’s transcriptional activity (Levy and Darnell, 2002; Wen et al., 1995; Shen et al., 2004), though the molecular mechanisms underlying Ser^{727} dephosphorylation have not yet been fully delineated. Though the serine/threonine phosphatases PP1 and PP2A have already been identified to dephosphorylate Ser^{727} (Lütticken et al., 1995; Woetmann et al., 1999; Togi et al., 2009; Haridas et al., 2009), the regulatory subunit that confers selectivity, specificity, and subcellular localization of PP1 towards STAT3 has so far and contrary to PP2A (Togi et al., 2009) remained unidentified. For the EGF pathway, we now demonstrate that Ser^{727} phosphorylation of STAT3 decreases on the condition that PP1b associates with activated SHOCA-2.

Dephosphorylation of Tyr^{705} constitutes another critical step in the events leading to STAT3 deactivation (Heinrich et al., 2003). One of the phosphatases involved in this event is the receptor protein tyrosine phosphatase delta (PTPRD), which has been shown to act as a tumor suppressor in both colon cancer and other neoplasms (Zhang et al., 2007; Veeriah et al., 2009). Though PP1b constitutes yet another phosphatase involved in STAT3 inactivation, it remains unknown whether it acts indirectly via a tyrosine-specific phosphatase (e.g. SHP (Neviani et al., 2005)), or directly through target promiscuity (MacKintosh et al., 1996; Shi, 2009). A dual specificity of PP1b's catalytic activity is not unparalleled as a comparable mechanism has been described for the protein tyrosine phosphatase SHP-2.
(Wu et al., 2002). Alternatively, the enhanced physical interaction between PP1b, SHOCA-2 and STAT3 following EGF binding to its receptor may be sufficient to inactivate STAT3 secondary to conformational changes independent of any catalytic activity (Lee et al., 2009). The precise molecular mechanism by which Tyr^{705} is de-phosphorylated remains to date unidentified. Irrespective of this lack of knowledge, inappropriate STAT3 regulation is the noticeable consequence of aberrant SHOCA-2 expression and results in an up-regulation of genes involved in cell proliferation.

The reduction of SHOCA-2 expression in the CRC cell line SW480 also produced phenotypic features characteristic of EMT. Such a transformation has been implicated in the conversion of early stage malignancies towards more aggressive tumors showing invasive growth and the formation of metastasis (Thiery et al., 2009). Consistent with a role in EMT, LOH at SH2D4A was mostly detected in patients with a disease stage typically associated with metastasis. It is thus conceivable that the loss of SHOCA-2 plays a part in tumor progression not only through its impact on tumor cell growth but possibly also via its contribution to EMT.

EGFR and STAT3 constitute promising drug targets in the treatment of cancer including CRC (Quesnelle et al., 2007). However, therapeutic responses are observed in only 10 to 20% of patients treated with EGFR antagonists (Ciardiello and Tortora, 2008) and none of the available STAT3 inhibitors can to date be considered as viable drug candidates because of their limited efficiency in disrupting STAT3 homodimerization (Yue and Turkson, 2009). Detailed knowledge about the regulatory network that controls STAT3 activity may likely provide novel rationale for the design of strategies to interfere with a STAT3-mediated activation in cancer cells.
Experimental Procedures

A detailed description of the methods used can be found in Supplemental Experimental Procedures.

Plasmids, antibodies, cells and reagents.
Antibodies, expression plasmids, cells and reagents were either gifts from investigators or bought commercially, as specified in Extended Experimental Procedures.

Bioinformatic analysis.
The SHOCA-1 (SH2D4B) and SHOCA-2 (SH2D4A) sequences were analyzed using NCBI, ENSEMBL (Hubbard et al., 2009), the kinBase database (http://www.kinase.com/kinbase/), the TCoffee software (Notredame et al., 2000), and Oncomine (Rhodes et al., 2007).

TAP purification and mass spectrometry.
The TAP purification (Chen and Gingras, 2007) and mass spectrometry (Hess et al., 2008) were carried out as previously reported.

Western Blot, immunoprecipitation and cell fractionation.
Cells lysates were either separated on a 8%-12% SDS-PAGE or immunoprecipitated with protein G Plus-Sepharose beads coupled to specified antibodies and analyzed by western blotting. Cell fractionation was achieved using the CelLytic Nuclear extraction kit (Sigma).

Luciferase reporter gene assays.
HeLa cells were transfected with a firefly luciferase reporter SIE (sis-inducible element) plasmid plus a Renilla luciferase plasmid (both Promega) and plasmids that encode either STAT3 or SHOCA-2. Cell extracts were tested using the Dual Luciferase Assay System (Promega) and analyzed using a luminometer. Firefly luciferase activity was normalized to Renilla luciferase activity.
Cell cycle analysis.
Cells were incubated for 15 min with ice-cold hypotonic Propidium Iodide staining solution and subsequently analyzed by flow cytometry (FACSCalibur, Becton Dickinson) using the FlowJo software (Tree Star, Oregon Corporation). G1, S and G2/M phases were defined using the mathematical Watson Pragmatic model.

Colony formation and anchorage-independent growth assays.
For colony formation assay, SW480 cells co-transfected with a control vector or plasmids that encode STAT3, SHOCA-2, SHOCA-2 PP1Mut2 or SHOCA-2 Y131A were cultured for 15 days under neomycin selection (G418: 1 mg/ml) and then analyzed for colony frequency following fixation and crystal violet staining of the cells. SW620 cells transfected with a control vector or plasmids that encode SHOCA-2, SHOCA-2 PP1Mut2 or SHOCA-2 Y131A, cultured under neomycin selection (G418, 1 mg/ml) in a soft agar (CytoSelect, Cell Biolabs) for 15 days and then analyzed for colony frequency.

In vivo analyses of tumor formation.
SW480 cells (5 × 10^6) were grafted to the flank of 6 week-old female BALB/c nude mice and the size of the tumor was measured weekly. Mice were sacrificed after 5 weeks and tumors were removed for further analysis. The animal experiments were carried out in accordance with the legal requirements of the Swiss veterinary authorities.

Human tissues and immunohistochemistry.
For the genetic and epigenetic study, specimens from 70 CRC patients were obtained from the Department of Surgery of the Kantonsspital Aarau, Switzerland and from the Department of Gastroenterology of the Inselspital in Berne, Switzerland, with approval of the local medical ethics boards and written consent from patients. Tissues were stored in RNAlater reagent at -70°C. Single-punch tumor samples from 501 patients (Institute of Pathology, Stadtspital Triemli, Zürich, Switzerland) were analyzed by tissue microarrays as described previously (Zlobec et al., 2010). SHOCA-2 expression and
STAT3 phosphorylation were quantified using the H score (Hirsch et al., 2003).

**Mutational screening, loss of heterozygosity and gene dosage analysis.**
The complete coding sequence including flanking intronic regions of the
*SH2D4A* gene was analyzed in all cancers lacking SHOCA-2 protein expression and/or that were positive for LOH. LOH for the *SH2D4A* locus was
determined using a set of 3 flanking microsatellite markers and 6 gene-specific single nucleotide polymorphisms (SNP) (for details see Extended
Experimental Procedures). Gene dosage was determined for each cancer
using a chromosome 8-specific multiplex ligation-dependent probe
amplification (MLPA) assay (SALSA P014-1A, MRC Holland) and data were
processed by the GeneMarker software package (Softgenetics).

**Methylation analysis.**
Genomic DNA was converted by bisulfite treatment using the EZ DNA
Methylation Kit™ according to the manufacturer’s instructions (Zymo
Research). CpG island regions were amplified by PCR from bisulfite treated
DNA and sequenced with the PyroMark Q24 pyrosequencing system (Qiagen)
allowing the quantitation of methylated CpG sites (PCR conditions and primer
sequences are given in Extended Experimental Procedures). Bisulfite treated
genomic DNA, previously *in vitro* methylated with M.SssI methyltransferase
(New England BioLabs), served as the positive control.

**Statistical analyses.**
Student’s t test, Likelihood ratio test, Chi-square test, Gehran-Wilcoxon test,
Fisher’s Exact test, binomial test for equal proportions, Log-rank test,
multivariate Cox regression analysis and Kaplan-Meier method were used for
statistical analyzes, as indicated.
Acknowledgements

We would like to thank Annick Peter, Dorothea Maass and Ragna Sack for expert technical help; Dr. Giancarlo Marra for microarray data; Dr. Fabrizio Bianchi for statistical analyzes; Dr. Bérengère Fayard and Dr. Laura Trinkle-Mulcahy for valuable discussions; Drs. Nancy Hynes, Thomas Barthlott, Poul Sørensen and Gerhard Christofori for critical reading of the manuscript; Sabrina Harris for secretarial assistance.

This work was supported by grants from the Swiss National Science Foundation (to G.A.H.), the Krebsliga beider Basel (to S.L. and G.A.H.) and the Oxford Biomedical Research Centre (to G.A.H.).

S.L. and G.A.H. designed the experiments, analyzed the data and composed together with P.S. the manuscript; S.L. performed experiments including Western Blot, immunoprecipitation, luciferase assays, mutagenesis, TAP assay, phosphatase assay, bioinformatic analyses, colony formation and anchorage-independent growth assays, and in vivo experiments; S.Z., K.H. and M.Ke. participated in the identification and cloning of SH2D4B and SH2D4A genes; D.H. performed the mass spectrometry analysis; supervised by P.S. and K.H., M.Ko. and S.W. performed the human genetic and epigenetic studies; immunohistochemistry was done by S.P.; K.T. provided human CRC samples; I.Z. performed statistical analyzes; L.T. supplied and analyzed CRC tissue microarray.

The authors declare no competing financial interests.
References


Lu, Y., Liang, K., Li, X., and Fan, Z. (2007). Responses of cancer cells with wild-type or tyrosine kinase domain-mutated epidermal growth factor receptor (EGFR) to EGFR-targeted therapy are linked to downregulation of hypoxia-inducible factor-1alpha. Mol Cancer 6, 63.


Figure legends

Figure 1. Reduced SHOCA-2 expression in human colorectal cancers correlates with advanced disease stage.

A) SHRDA

B) SHOC1A

C) SHOC1B

D) SHOC-1 FRET

SHOCA-1

SHOCA-2

Degree of conservation

Low

High

Frequency of tumors with low and high SHOCA-2 expression (%)

Stage

I

II

III

IV

H score<100 (n=38)

H score 100-200 (n=67)

H score>200 (n=275)

P=0.0119

Survival probability

Time (months)

0

25

50

75

100

125

150

150
(A) Left, location of SHOCA-2 on human chromosome (huChr) 8p. Middle, degree of conservation between the two human SHOCA sequences with higher degrees of conservation specified by darker shades of grey; box, a consensus MyPhoNE sequence typically present in myosin homologues (Hendrickx et al., 2009). Right, predicted domain organization of SHOCA; the amino acid identity between SHOCA-1 and -2 is given for the MyPhoNE and SH2 domains. Of note, SHOCA-2 protein contains only one coiled-coil domain (CC), namely CC2.

(B) Immunohistochemical analysis of a single tissue microarray containing 400 samples from unselected, untreated patients with sporadic CRC. SHOCA-2 expression was quantified using the H-score that determines the percentage of positive tumor cells multiplied by their staining intensity generating individual groups with scores of >200 (n=275), defining a strong expression (green); 100-200 (n=87), identifying a moderate expression (blue); <100 (n=38), characterizing a low or absent expression (red).

(C) Representative immunohistochemical analysis of tumor tissue of distinct CRC stages depicting either diffuse (>90% cells stained; left), patchy (<30%; middle) or no SHOCA-2 expression (right). T = tumor tissue, N = normal tissue. Scale bar, 200 mm.

(D) Kaplan-Meier survival analysis of patients with high (green), moderate (blue) and low SHOCA-2 expression (red) for which sufficient clinical information was available (log-rank test p = 0.0118 when comparing high to low SHOCA-2 expression).

See also Figure S1 and Table S1
Figure 2. Genetic and epigenetic silencing of SH2D4A encoding SHOCA-2.

A

B

C

D

E

Tumor stage

<table>
<thead>
<tr>
<th>Patient N°</th>
<th>I-II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17-18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19-20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23-25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26-27</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(A) HuChr8p copy number variation profiles in SHOCA-2-positive and SHOCA-2-negative CRC. Values <0.75 indicate chromosomal deletions (Del) whereas values >1.25 designate chromosomal duplications (Dup). Red dashed lines, threshold.

(B) Top, comparison of the SH2D4A c.342-5T>C (rs17128221) SNP frequency between 83 unrelated healthy Caucasian controls and 16 CRC patients (Chi square p value = 0.006). Bottom, SNP rs17128221 RT-PCR (right) and sequencing analysis (left), the diagram depicts exon usage and the arrows indicate the annealing sites for the primers.

(C) i: top, alignment of amino acid sequences for the SH2 domain (AA 320-454) of human SHOCA-2 (AAH82982) with homologues of mus musculus (AAI16683), rattus norvegicus (NP_001012048), and gallus gallus (XP_420452); bottom, sequence analysis of patient tumors. ii: immunoblot analysis of HeLa cells transfected to express Flag-fusion proteins of the indicated human SHOCA-2 mutations.

(D) DNA methylation analysis of CpG island located in the 5'UTR of SH2D4A comparing normal and tumor tissue.

(E) Summary of all genetic and epigenetic changes detected in a cohort of 27 informative CRC samples; CNV = copy number variation, SNP = single nucleotide polymorphism. Green = normal finding, red = pathological result, n.d. = not done, roman numerals = number of alleles affected.

See also Figure S2
Figure 3. SHOCA-2 is phosphorylated following EGFR activation and inhibits STAT3 activity.

A

B

C

D

E

Figure 3

SHOCA-2-Flag

Figures A, B, C, D, and E show the effects of SHOCA-2 on STAT3 activity and phosphorylation following EGFR activation.
(A) SHOCA-2-Flag expressing, serum-starved HeLa cells were stimulated as indicated with EGF (100 ng/ml). Cell lysates were immunoprecipitated with anti-phosphotyrosine (pTyr) antibody or rabbit IgG and immunoblotted to detect Flag and pTyr.

(B) Serum-starved HeLa cells were transiently transfected with either a control vector or a vector encoding the human SHOCA-2-Flag and stimulated as indicated with EGF (100 ng/ml); immunoblotting for the detection of the indicated proteins. The ratio of phosphoprotein to total protein (designated P/Total) was determined by densitometry.

(C) Serum-starved HeLa cells that express normal (control, 100%) or 10% of SHOCA-2 (SHOCA-2 KD) were stimulated for 5, 10 and 15 min with EGF (100 ng/ml). The indicated proteins were detected by immunoblotting. Densitometry was used to determine the ratio of phosphoprotein to total protein (designated P/Total).

(D) Serum-starved HeLa cells and SHOCA-2 KD HeLa cells transiently co-transfected with a SIE firefly luciferase reporter, a renilla luciferase plasmid and a SHOCA-2 expression plasmid were stimulated for 6 hrs with EGF (100 ng/ml; filled bar) or left untreated (open bar). Luciferase assay was performed on lysates. Renilla luciferase activity was used to normalize transfection efficiency. Data represent the mean ± SD of three independent, triplicate assays (Student’s t test, **p < 0.01; ***p < 0.005).

(E) Immunoblotting of HeLa, PAE and MEF lysates immunoprecipitated with anti-EGFR, anti-SHOCA-2, anti-Flag or mouse IgG antibodies for the detection of EGFR, STAT3, SHOCA-2 and Flag.

See also Figure S3
Figure 4. SHOCA-2 controls STAT3 activity in a PP1b-dependent fashion.

A

B

C

D

E

F

111
(A) Lysates from HEK293 cells (bottom) and HEK293 cells expressing SHOCA-2-Flag (top) were immunoprecipitated with anti-PP1b antibody or IgG and immunoblotted for PP1b, and Flag (top) or SHOCA-2 (bottom). Human SHOCA-2 is detected as isoforms due to exon 5 skipping.

(B) Lysates from HEK293 cells expressing CTAP fusion proteins and EGFP-tagged PP1 isoforms were immunoprecipitated with IgG sepharose beads (recognizing CTAP fusion proteins) and immunoblotted for the detection of CTAP and EGFP fusion proteins.

(C) Left, diagram representing wild-type and mutant SHOCA altered PP1b-binding sites (boxes) are marked by crosses. Middle, CTAP fusion proteins and PP1b were detected in total and immunoprecipitated lysates from transfected HEK293 cells. Right, SHOCA-2, Flag and PP1b, respectively, were detected in total lysates and anti-Flag immunoprecipitates from HEK293 cells expressing wild type or mutant SHOCA fusion proteins.

(D) Left, serum-starved HeLa cell expressing the human SHOCA-2-Flag fusion protein were stimulated with EGF (100 ng/ml) for the indicated times. Lysates were analyzed unmanipulated or were immunoprecipitated with anti-EGFR antibody for the detection of EGFR, STAT3, Flag and PP1b by immunoblotting. Right, ratio of immunoprecipitated protein to total protein (designated IP/Total) as determined by densitometry.

(E) Serum-starved HeLa cell co-expressing the indicated proteins were stimulated with EGF (100 ng/ml) for the specified times. Lysates were immunoblotted for the detection of phospho-EGFR (Tyr845), phospho-STAT3 (Tyr705/Ser727), STAT3, Flag, and GAPDH.

(F) Serum-starved HeLa cells transiently co-transfected with a SIE firefly luciferase reporter, a renilla luciferase plasmid and plasmids encoding either STAT3, the human SHOCA-2-Flag fusion protein or a mutant human SHOCA-2-Flag fusion protein unable to associate with PP1b (PP1Mut2) were stimulated for 6 hrs with EGF (100 ng/ml; filled bar) or left untreated (open bar). Luciferase assay was performed on lysates. Renilla luciferase activity was used to normalize transfection efficiency. Data represent the mean ± SD of three independent, triplicate assays (Student's t test, *p < 0.05; **p < 0.01; ***p < 0.005).

See also Figure S4 and Table S2
Figure 5. SHOCA-2 and PP1b arrest tumor cell proliferation sustained by STAT3.

Figure 5

A

<table>
<thead>
<tr>
<th>shRNA</th>
<th>Control</th>
<th>SHOCA-2 KD</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGF</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>PTyr705 STAT3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAT3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c-Myc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jun B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclin D1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHOCA-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAPDH</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

- Stage I-II
- Stage III-IV

Frequency of positive tumor cells (%)

C

- Number of colonies

D

- Overexpression

E

- Number of colonies
(A) Left panel: Serum-starved HeLa cells that expressed wild type (control) or reduced (SHOCA-2 KD) SHOCA-2 levels were either left untreated or stimulated with EGF (100 ng/ml) for 6 and 12 hrs. Cell lysates were immunoblotted for the detection of phospho-STAT3 (Tyr705), STAT3, c-Myc, Jun B, Cyclin D1, SHOCA-2 and GAPDH. Right panel: After 24 hrs of serum starvation, control and SHOCA-2 KD HeLa cells were stimulated with EGF (100 ng/ml) for 12 hrs. The cell cycle profile was evaluated by flow cytometry using propidium iodide staining (PI). The bar graph shows the percentage of cells in the indicated phases of the cell cycle. The data is representative of 3 separate experiments, each performed in triplicate (*p < 0.05; ***p < 0.005 using Student's t test).

(B) CRC tissue sections from a cohort of 27 informative CRC patients (see Figure 2E) were analyzed for the proportion of tumor cells that stained for SHOCA-2 and Ki-67 (Wilcoxon test, *p < 0.05; ***p < 0.005).

(C) SW480 cells transfected to express the indicated empty (pYN3218, Flag) and recombinant vectors were grown for 15 days under neomycin selection and then scored for the number (top) and appearance of colonies (bottom). The graph is representative of two independent experiments performed in triplicates (Student's t test, *p < 0.05; **p < 0.01; ***p < 0.005).

(D) SW480 cells were transiently transfected with a control vector or recombinant vectors as indicated, serum-starved, and then stimulated with EGF (100 ng/ml) for 10 and 15 min. Lysates were immunoblotted to detect phospho-STAT3 (Tyr705 and Ser727), STAT3, phospho-Akt (Ser473), Akt, phospho-ERK1/2 (Thr202/Tyr204), Flag and GAPDH.

(E) SW620 cells expressing wild-type human SHOCA-2 or a SHOCA-2 mutant (PP1\textsubscript{Mut2}, Y131A) were grown for 15 days under neomycin selection in soft agar and then scored for the number (top) and appearance of colonies (bottom). The graph is representative of two independent experiments performed in triplicates (Student's t test, ***p < 0.005).

See also Figure S5
Figure 6. SHOCA-2 is required for maintaining epithelial morphology and its down-regulation promotes tumorigenesis in vivo.
(A) Phase-contrast images of control SHOCA-2 expressing SW480 cells and polyclonal SHOCA-2 knockdown (KD) SW480 cells grown at different densities. Scale bar, 100 mm.

(B) Immunoblot analysis of E-cadherin, N-cadherin, Vimentin, ZEB-1, SNAIL, SHOCA-2 and GAPDH protein expressions in lysates from SW480 stable transfectants (shRNA control, SHOCA-2 KD).

(C) Proliferation of SW480 cells that have been stably transfected with shRNA for the knock-down of SHOCA-2. The cells have been grown in the presence or absence of EGFR inhibitor (AG1478). Data represent the mean ± SEM of two independent experiments.

(D) SW480 cells with or without a knock-down of SHOCA-2 were transfected to express either a control plasmid (designated Flag) or a murine (m) SHOCA-2-Flag fusion protein that was not targeted by the shRNA used. Transfected cells selected with neomycin and scored for the number of colonies detected (top). Cell lysates were immunoblotted for SHOCA-2 and GAPDH protein expression (bottom). The data are representative of two independent experiments.

(E) Monitoring tumor incidence and size in nude mice (n=5 mice per group) grafted with either SHOCA-2 wild type (shRNA CT) or knock-down SW480 cells (SHOCA-2 KD). Data represent the mean ± SEM of two independent experiments.

(F) Immunohistochemistry of grafted SW480 tumor cells either wild type or knock-down for SHOCA-2 expression (upper left panel; scale bar, 100 mm) and of human CRC tissues (right; scale bar, 50 mm). Left upper panel: Immunohistochemical analysis of in vivo grafted SW480 cells for STAT3 phosphorylated at Tyr705; Left lower panel: H-score analysis for Tyr$^{705}$ and Ser$^{727}$ STAT3 phosphorylation in CRC samples with low to absent SHOCA-2 expression, displayed as a relative percentage of the biopsies investigated. Right: analysis of CRC tissues on consecutive section for the detection of SHOCA-2 and STAT3 phosphorylated either at Tyr$^{705}$ or Ser$^{727}$. See also Figure S6
Supplementary Information

The 8p21.3 encoded SHOCA-2 represses STAT3 activation and acts as a tumor suppressor in colorectal cancer

Sébastien Loeffler, Michal Kovac, Stefan Weis, Salvatore Piscuoglio, Saulius Zuklys, Marcel Keller, Katrin Hafen, Daniel Hess, Kaspar Truninger, Inti Zlobec, Luigi Terracciano, Karl Heinimann, Primo Schär and Georg A. Holländer

Inventory of Supplemental Information

Figure S1, related to Figure 1
Table S1, related to Figure 1
Figure S2, related to Figure 2
Figure S3, related to Figure 3
Figure S4, related to Figure 4
Table S2, related to Figure 4
Figure S5, related to Figure 5
Figure S6, related to Figure 6
Supplemental Experimental Procedures
Supplemental References
Supplementary Figure 1
Supplementary Data

A

B

C

D

E

F

118
Figure S1. Related to Figure 1. SHOCA sequence analyses and expression profiling in mouse tissue and in tumors. (A) The figure represents an unrooted evolutionary tree analysis of SH2D4 paralogues using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) method. Concurrent with the development of vertebrates (thick line), two distinct SH2D4 family members developed. (B) Tissues from 10 week-old Balb/c mice were analyzed for SH2D4B and SH2D4A expression using real time RT-PCR. The results are normalized to 18S rRNA levels. The black bars correspond to SH2D4B mRNA levels and the gray bars correspond to SH2D4A mRNA levels. (C) The N-terminal amino acid sequences (AA1-51) of human SHOCA-1 and human SHOCA-2 are aligned with human myosin homologues: MYO16 (Myosin-XVI), MYPT1 (Myosin phosphatase-targeting subunit 1, a.k.a. PPP1R12A), MYPT2 (Myosin phosphatase-targeting subunit 2, a.k.a. PPP1R12B), PPP1R12C (Protein phosphatase 1 regulatory subunit 12C), MYPT3 (Myosin phosphatase-targeting subunit 3, a.k.a. PPP1R16A) and ANKHD4 (Ankyrin repeat domain-containing protein 4, a.k.a. PPP1R16B). The box identifies the MyPhone consensus sequence typically present in myosin homologues. (D) From a cohort of 571 unselected patients with sporadic CRC, paraffin-embedded tissue for tissue microarrays (TMA) was available for 501 patients. SHOCA-2 expression analysis was carried out on 400 samples with sufficient tumor tissue (>50% biopsy) and documented disease stage. Given an informative immunohistochemical analysis and clinical data, survival probability could be calculated for 300 of these patients. LOH was investigated on fresh frozen tissues for 70 of the 571 patients, and epigenetic and genetic analyses were carried out on 27 cases informative for LOH. (E) SHOCA-2 mRNA levels were analyzed by using 9 public gene expression datasets available for meta-analysis (Oncomine)'. n designates the number of cases included in the analysis. The Y-axis represents expression units that have been normalized as described elsewhere (http://www.compendiabio.com/support/normal.htm). Shaded boxes represent the interquartile range (i.e. the 25th-75th percentile); whiskers define the 10th-90th percent range; bars indicate the median value; and closed circles identify outliers. The p-value was calculated using Student's t-Test. (F) Tissues from healthy human mucosa (n=4) and tumors (n=10) were analyzed for SHOCA-1 expression using real time RT-PCR. The results are normalized to GAPDH levels. The dots correspond to the relative SHOCA-1 expression of each sample and the bars represent the group's median value. The p-value was calculated using Mann-Whitney test.
### Supplementary Table 1

#### A

<table>
<thead>
<tr>
<th>H-score</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;100</td>
<td>Stage I vs II &lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>Stage I vs III 0.0613</td>
</tr>
<tr>
<td></td>
<td>Stage I vs IV 0.0588</td>
</tr>
<tr>
<td></td>
<td>Stage II vs III 0.3692</td>
</tr>
<tr>
<td></td>
<td>Stage II vs IV 0.0143</td>
</tr>
<tr>
<td></td>
<td>Stage III vs IV 0.1083</td>
</tr>
<tr>
<td>100-200</td>
<td>Stage I vs II 0.0039</td>
</tr>
<tr>
<td></td>
<td>Stage I vs III &lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>Stage I vs IV 0.0003</td>
</tr>
<tr>
<td></td>
<td>Stage II vs III 0.1025</td>
</tr>
<tr>
<td></td>
<td>Stage II vs IV 0.3865</td>
</tr>
<tr>
<td></td>
<td>Stage III vs IV 0.4386</td>
</tr>
<tr>
<td>&gt;200</td>
<td>Stage I vs II &lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>Stage I vs III 0.9005</td>
</tr>
<tr>
<td></td>
<td>Stage I vs IV 0.7532</td>
</tr>
<tr>
<td></td>
<td>Stage II vs III 0.4610</td>
</tr>
<tr>
<td></td>
<td>Stage II vs IV &lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>Stage III vs IV 0.9022</td>
</tr>
</tbody>
</table>

#### B

<table>
<thead>
<tr>
<th>H-score</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>100-200 vs &gt;200</td>
<td>Stage I vs II 0.2628</td>
</tr>
<tr>
<td></td>
<td>Stage I vs III 0.0189</td>
</tr>
<tr>
<td></td>
<td>Stage I vs IV 0.0009</td>
</tr>
<tr>
<td></td>
<td>Stage II vs III 0.0739</td>
</tr>
<tr>
<td></td>
<td>Stage II vs IV 0.062</td>
</tr>
<tr>
<td></td>
<td>Stage III vs IV 0.1298</td>
</tr>
<tr>
<td>&lt;100 vs &gt;200</td>
<td>Stage I vs II 0.0139</td>
</tr>
<tr>
<td></td>
<td>Stage I vs III 0.0368</td>
</tr>
<tr>
<td></td>
<td>Stage I vs IV 0.112</td>
</tr>
<tr>
<td></td>
<td>Stage II vs III 0.5808</td>
</tr>
<tr>
<td></td>
<td>Stage II vs IV 0.5409</td>
</tr>
<tr>
<td></td>
<td>Stage III vs IV 0.8622</td>
</tr>
</tbody>
</table>

#### C

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Log-rank (p-value)</th>
<th>Wilcoxon (p-value)</th>
<th>Cox (p-value)</th>
<th>HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100-200 vs &gt;200</td>
<td>0.2312</td>
<td>0.2154</td>
<td>0.2365</td>
<td>0.81 (0.57-1.15)</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I (no events)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Stage II</td>
<td>0.146</td>
<td>0.096</td>
<td>0.156</td>
<td>0.67 (0.38-1.17)</td>
</tr>
<tr>
<td>Stage III</td>
<td>0.582</td>
<td>0.75</td>
<td>0.586</td>
<td>0.82 (0.39-1.7)</td>
</tr>
<tr>
<td>Stage IV</td>
<td>0.843</td>
<td>0.824</td>
<td>0.846</td>
<td>0.94 (0.51-1.74)</td>
</tr>
<tr>
<td>H-scores</td>
<td>&lt;100 vs &gt;200</td>
<td>0.0118</td>
<td>0.0133</td>
<td>0.0134</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I (no events)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Stage II</td>
<td>0.355</td>
<td>0.553</td>
<td>0.36</td>
<td>0.62 (0.22-1.73)</td>
</tr>
<tr>
<td>Stage III</td>
<td>0.548</td>
<td>0.551</td>
<td>0.55</td>
<td>0.78 (0.34-1.77)</td>
</tr>
<tr>
<td>Stage IV</td>
<td>0.569</td>
<td>0.809</td>
<td>0.58</td>
<td>0.82 (0.41-1.64)</td>
</tr>
</tbody>
</table>
Table S1. Related to Figure 1. Statistical analyses for association of SHOCA-2 H-score, stage and survival time (A) Differences in the stage distribution among each H-score category (binomial test for equal proportions). (B) Unadjusted p-values for association of H-scores and stage (Chi-Square or Fisher’s Exact test). (C) P-values for survival time differences overall and stratified by stage using three different statistical tests.
Figure S2. Related to Figure 2. Epigenetic and genetic analyses on SH2D4A gene. (A) Loss of Heterozygosity (LOH) analysis for SH2D4A. Loss of heterozygosity analysis for the SH2D4A locus. Panels i and ii depict 3 flanking and 1 gene-specific microsatellite marker, respectively, that cover the region (approximately 7Mb) from 8p21.3 to the telomeric boundary of 8p22. Panel iii demonstrates for cancer sample number 17 the analysis of 5 gene-specific SNPs and reveals a selective loss of a single SHOCA-2 allele. (B) Left, Sp1 binding affinity differs between methylated and unmethylated SH2D4A DNA oligomers. The Sp1 specific gel-shifts were carried out using a fluorescein-labeled oligomer binding Sp1 (CT+). CT+ and SW48 nuclear extract were incubated in a competitive manner in the absence or presence of increasing concentrations of an unlabeled oligomers. This oligomer sequence, which represents part of the genomic SH2D4A sequence, had either an unmethylated (Unmeth) or a methylated (Meth) Sp1 binding site. As a negative control (designated Oligo), CT+ oligomer was incubated in the absence of nuclear extract. Data from four independent experiments are shown and represent the mean and the interquartile range (i.e. the 25th-75th percentile) of the percentage of the fluorescence signal obtained with experimental samples in comparison to control (CT+; *p < 0.05 using Mann-Whitney test). The presence of Sp1 in the shifted bands was confirmed by Western blotting (WB). Right, HeLa cells were transiently transfected with either a control vector or a vector encoding Sp1. Immunoblotting for the detection of the indicated proteins.
Supplementary Figure 3

A  Overexpression

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>SHOCA-2-Flag</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EGF</strong></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P(Ser473) AKT</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>AKT</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Ratio P/Total</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>P ERK 1/2</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>ERK 1/2</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Ratio P/Total</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>SHOCA-2-Flag</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

B  Overexpression

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>SHOCA-2-Flag</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EGF</strong></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P(Ser660) PKC</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>GAPDH</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Ratio P/Total</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>SHOCA-2-Flag</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

C  shRNA

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>SHOCA-2 KD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EGF</strong></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P(Ser473) Akt</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Akt</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Ratio P/Total</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>P ERK 1/2</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>ERK 1/2</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Ratio P/Total</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>SHOCA 2</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>GAPDH</td>
<td>0.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

D

Relative abundance activity

E

EGF

MEFs

STAT3

F

<table>
<thead>
<tr>
<th>IP</th>
<th>Flag</th>
<th>Total lysate</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHOCA-2-Flag</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SHOCA-2-SH2Δ-Flag</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>STAT3</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

G

<table>
<thead>
<tr>
<th>IP</th>
<th>STAT3-HA</th>
<th>Shoca-1-CTAP</th>
<th>EGFP-CTAP</th>
<th>Total lysate</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT3-HA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure S3. Related to Figure 3. Effect of SHOCA-2 on EGFR-mediated signaling. (A,B) Serum-starved HeLa cells were transiently transfected with either a control vector or a vector encoding the human SHOCA-2-Flag sequence and were then stimulated as indicated with EGF (100 ng/ml); immunoblotting for the detection of the indicated proteins. The ratio of phosphoprotein to total protein (designated P/Total) was determined by densitometry. (C) HeLa cells were stably transfected either with a plasmid encoding an shRNA specific for SHOCA-2 or with a plasmid devoid of an insert. Following serum starvation for 24 hrs, the cells were stimulated with EGF (100 ng/ml) as indicated. Western blotting for the indicated proteins were performed on aliquots of transfected cell lysates. The efficiency of SHOCA-2 silencing is indicated as a percentage of control transfected cells. Densitometry was used to determine the ratio of phosphoprotein to total protein signal (designated P/Total). (D) A SIE firefly luciferase reporter construct and a renilla luciferase plasmid were transiently transfected together with a STAT3 expression vector into unmanipulated HeLa cells. Transfected cells were serum-starved for 24 hrs and then either stimulated with EGF (100 ng/ml, filled bar) for 6 hrs or left untreated (open bar). The renilla luciferase activity was used to normalize for transfection efficiency. Data are representative of 3 experiments, each performed in triplicate (*p < 0.05; **p < 0.01; ***p < 0.005 using Student’s t test). (E) Immunoblotting of lysates from PAE cells, HeLa, and wild type and Stat3 deficient MEFs for the detection of EGFR and Stat3, respectively. (F) HeLa cells were transfected to express a Flag-fusion protein of either the wild-type form of human SHOCA-2 or a mutant form of human SHOCA-2 that lacks the SH2 domain (designated SHOCA-2-SH2Δ). Cell lysates were immunoprecipitated with anti-Flag antibodies and subsequently analyzed by immunoblotting using antibodies specific for Flag, Stat3, and EGFR (left panel). Total lysates were used as loading controls (right panel). (G) HEK293 cells were transfected to express STAT3-HA and a C-terminal TAP-tag fusion proteins of SHOCA-1 or EGFP. Cell lysates were immunoprecipitated with IgG beads, and immunoblotted with either antibodies specific for HA-tag or with Rabbit IgG recognizing the protein A component of the TAP-tag.
Figure S4. Related to Figure 4. Interaction of SHOCA-2 with the EGFR/STAT3/PP1β complex. (A) Multiple reaction monitoring (MRM) comparison of two selected peptides from SHOCA-1, PP1β and HSP71, respectively. Red and blue lines represent peptides retrieved from the SHOCA-1-CTAP and control EGFP-CTAP pull-downs, respectively. The transition signals are labeled with the mass (m/z) of the peptides. Peptides derived from non-specific interactions (e.g., heat shock protein HSP71) were identified in both eluates whereas SHOCA-1 and PP1β peptides were only found in the SHOCA-1-CTAP immunoprecipitation. (B) Cell lysates from HEK293 cells transfected to express SHOCA-1-Flag fusion protein were immunoprecipitated either with an antibody specific for PP1β or with goat IgG and then immunoblotted with antibodies specific for Flag and PP1β. (C) HEK293 cells were transfected to express the indicated TAP-tag fusion proteins. The cell lysates were immunoprecipitated with IgG beads, and immunoblotted with either antibody specific for PP1β (loading control) or with Rabbit IgG recognizing the protein A component of the TAP-tag. (D) Top, HeLa cells expressing an EGFP-PP1β fusion protein were stained with anti-SHOCA-2 antibody and DAPI to detect colocalization by confocal microscopy. Bottom, endogenous SHOCA-2 was detected in cytoplasmic (GAPDH+) and nuclear (Histone H1+) HeLa cell lysates immunoprecipitated with anti-PP1β. (E) HEK293 cells were transfected to express either a wild-type SHOCA-1-Flag fusion protein or one of two deletion mutants AA1-80 or AA51-94, fused to Flag (left panel; the alleged PP1-binding sites are indicated as vertical bars). Lysates from transfected cells were immunoprecipitated with anti-Flag antibodies and immunoblotted with antibodies directed against Flag or PP1β (right panel). Interaction of wild type and mutant SHOCA-2 with EGFR and STAT3 (F) Serum-starved HeLa cell expressing the human SHOCA-2-Flag fusion protein were stimulated with EGF (100 ng/ml) for the indicated times. Lysates were immunoprecipitated with anti-Flag antibody, EGFR, STAT3, c-SRC, Flag, PP1β were detected by immunoblotting. (G) Lysates of HeLa cells expressing HA-tagged wild-type (WT) STAT3 or the Y705F STAT3 mutant and wild-type (WT) EGFR or the Y1069/1086F EGFR mutant were immunoprecipitated with an EGFR specific antibody. EGFR and HA were detected by immunoblotting. (H) HeLa cells were transfected with an expression plasmid encoding a Flag-fusion either wild-type (WT), PP1β or Y131A mutant protein of human SHOCA-2. Cell lysates were immunoprecipitated with anti-Flag antibodies and subsequently analyzed by immunoblotting using antibodies specific for Flag, STAT3, EGFR and PP1β (left panel). Total lysates were used as loading controls (right panel). (I) Serum-starved HeLa cells expressing human wild-type SHOCA-2-Flag or the Y131A mutant were stimulated with EGF (100 ng/ml) for the indicated times. Lysates were immunoprecipitated with a phosphotyrosine (pTyr) specific antibody. STAT3, Flag and pTyr were detected by immunoblotting.
## Supplementary Table 2

<table>
<thead>
<tr>
<th>Uniprot</th>
<th>Description</th>
<th>Shoca1-CTAP</th>
<th>EGFP-CTAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>P07437</td>
<td>Tubulin beta chain</td>
<td>49671</td>
<td>360</td>
</tr>
<tr>
<td>B2RS4Q7</td>
<td>SH2 domain containing 4B (Shoca1)</td>
<td>51120</td>
<td>0</td>
</tr>
<tr>
<td>P62140</td>
<td>Serine/threonine-protein phosphatase PP1-beta catalytic subunit</td>
<td>37187</td>
<td>0</td>
</tr>
<tr>
<td>P81947</td>
<td>Tubulin alpha-1B chain</td>
<td>50152</td>
<td>265</td>
</tr>
<tr>
<td>P08107</td>
<td>Heat shock 70 kDa protein 1</td>
<td>70052</td>
<td>88</td>
</tr>
<tr>
<td>P04517</td>
<td>Genome polyprotein</td>
<td>346164</td>
<td>182</td>
</tr>
<tr>
<td>P68104</td>
<td>Elongation factor 1-alpha</td>
<td>50141</td>
<td>81</td>
</tr>
<tr>
<td>P08238</td>
<td>Heat shock protein HSP 90-beta</td>
<td>83264</td>
<td>168</td>
</tr>
<tr>
<td>P62158</td>
<td>Calmodulin</td>
<td>16838</td>
<td>266</td>
</tr>
<tr>
<td>Q9H1R3</td>
<td>Myosin light chain kinase 2</td>
<td>64685</td>
<td>181</td>
</tr>
<tr>
<td>P42212</td>
<td>Green fluorescent protein</td>
<td>28886</td>
<td>357</td>
</tr>
</tbody>
</table>

Table S2. Related to Figure 4. LC-MS/MS analysis: comparison of peptides eluted following SHOCA-1- and EGFP-TAP purification. Proteins were identified using 2 or more unique peptides with a MASCOT score above 20. To distinguish between correct and incorrect peptide assignments, protein spectra with less than 4 peptides were manually evaluated. Spectral count: number of spectra identified for each protein.
Figure S5. Related to Figure 5. SHOCA-2 influences cell cycle progression and proliferation in STAT3-dependent fashion. (A) mRNA expression of SHOCA-2, Cyclin D1, Jun B, c-Myc, was assessed in 32 adenoma and 13 sporadic colorectal cancers by means of Affymetrix GeneChip U133 expression arrays. Shaded boxes represent interquartile range (i.e. the 25th-75th percentile), whiskers indicate the 10th-90th percent range, bars provide the median value. (B) H1975 cells were transfected to express either a vector without insertion (designated Flag) or a vector encoding Flag-fusion proteins of either wild type SHOCA-2, a SHOCA-2 mutant unable to bind to PP1δ (SHOCA-2-PP1δmut), or a SHOCA-2 mutant in which the tyrosine at position 131 was exchanged for an alanine (SHOCA-2-Y131A), or wild type Shoca-1. The cells were grown for 15 days under neomycin selection and then scored for the number (bar graphs) and appearance of colonies (microphotographs). The data is representative of two independent experiments performed in triplicates (Student’s t test, *p < 0.05; **p < 0.01; ***p < 0.005). (C) Proliferation of SW480 cells that have been stably transfected with a control (shRNA control) vector or a recombinant vector to knock-down SHOCA-2 expression (SHOCA-2 KD). Successfully transduced cells were grown either in the presence or in the absence of the STAT3 inhibitor S3I-201 diluted in DMSC (100nM). Data represent the
mean ± SD of two independent experiments carried out in triplicates. (D) SHOCA·2 protein expression in the CRC cell lines SW480 (primary tumor) and SW620 (metastasis from the same patient) was assessed using Western-blotting.
Figure S6. Related to Figure 6. SHOCA-2 knock-down SW480 cells display features of EMT. (A) Low power view of SW480 cells expressing either a control vector (shRNA control, left panel) or a vector to knock-down SHOCA-2 expression (SHOCA-2 KD, right panel). Scale bar, 100 μm. (B) Left, phase-contrast images of Shoca-2 KD SW480 cells grown in the presence or absence of EGFR inhibitor AG1478 diluted in DMSO (10μM; scale bar 50 μm). Right, immunoblot analysis of E-cadherin, Vimentin, Shoca-2 and GAPDH protein expressions in lysates from shRNA control and Shoca-2 KD SW480 cells cultured in the in the presence or absence of EGFR inhibitor AG1478 diluted in DMSO (10μM).
Supplemental Experimental Procedures

Plasmids, antibodies and reagents

Mouse and human SHOCA-1 and SHOCA-2 sequences were amplified from thymus cDNA libraries by PCR, subcloned into the expression vectors pcDNA-Flag 3.1, CTAP and NTAP (a kind gift from Dr. A.C. Gingras, Samuel Lunenfeld Research Institute at Mount Sinai Hospital, Toronto, Canada) and sequenced. Site-directed mutagenesis (Qiagen) was used to introduce specific mutations. The human EGFP-PP1α, β, γ1 constructs were generously provided by Dr. L. Trinkle-Mulcahy (Wellcome Trust Centre for Gene Regulation and Expression, University of Dundee, Scotland, UK), the EGFP-CTAP plasmid was a gift from Dr. S. Meloche (Institute of Research in Immunology and Cancer, Université de Montréal, Canada), the SIE firefly luciferase reporter plasmid was kindly made available by Dr. I. Behrmann (Life Sciences Research Unit, University of Luxembourg, Luxembourg), the EGFR wild-type and Y1068/1086F mutant constructs were generously provided by Dr. S. Sigismund (IFOM, Milano, Italy) and the STAT3 expression vector was obtained courtesy of Dr. J.Y. Yoo (Pohang University of Science and Technology, Pohang, Republic of Korea). Antibodies used were directed against PP2A (Upstate); Flag, ERK ½, anti-vimentin (all from Sigma); Histone H1, Jun B, c-Myc, cyclin D1, c-SRC and ZEB-1 (all from Santa-Cruz); GAPDH, SNAIL and PP1γ (all from Abcam); SH2D4A (Abnova); EGF receptor, Tyr845-phosphorylated EGFR, Ser660-phosphorylated PKC, phospho-tyrosine (designated pTyr), Tyr705/Ser727-phosphorylated STAT3, Akt, Ser473-phospho-Akt, Thr202/Tyr204-phospho-ERK 1/2 (all from Cell Signaling); GFP (Roche); anti-N-cadherin (Takara Biomedicalcs); anti-E-cadherin and STAT3 (both from BD Biosciences). EGF was obtained from Peprotech and used at
a concentration of 100 ng/ml. AG1478 was purchased from Axxora and the STAT3 inhibitor (S3I-201) was obtained from Merck.

Cell culture and transfection

HEK293, HeLa, SW480, SW620 cells (a kind gift from Dr. G. Marra, Institute of Molecular Cancer Research, University of Zurich, Switzerland), Murine embryonic fibroblasts (MEF) either wild type (fl/fl) or deficient for Stat3 (MEFΔ/Δ), a generous gifts of Dr. Valeria Poli, University of Turin, Torino, Italy, were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with Glutamax (4 mM), and 10% fetal calf serum (Invitrogen-Gibco). Porcine aortic endothelial (PAE) cells (kindly made available by Dr. P. Berger, Paul Scherrer Institut, Villigen, Switzerland) were grown in F-12 Nutrient Mixture (Ham) supplemented with Glutamax (4 mM), and 10% fetal calf serum (Invitrogen-Gibco). Transient transfections were accomplished using Lipofectamine 2000 (Invitrogen) for HEK293 and Eugene HD (Roche) for HeLa, SW480, SW620 cells according to the manufacturer's recommendations. For stable pShoca-1-CTAP or pEGFP-CTAP transfectants, HEK293 cells were co-transfected with the puromycin resistance plasmid, pBabe-puro (Cell Biolabs). Cells were first plated onto 100-mm plates and allowed to grow to 80% confluence, then rinsed twice with phosphate-buffered saline (PBS) and finally exposed to the plasmids/Lipofectamine-DMEM mixture. After overnight incubation, the cells were washed, split into a 150-mm plate and for selection grown in DMEM containing puromycin (10.8 μg/ml for the first 48 hrs and then 3.6 μg/ml). Several puromycin resistant clones were established. Cells expressing comparable amounts of either Shoca-1-CTAP or EGFP-CTAP protein (as analyzed by western-blotting) were selected for further analysis. For the stable knock-down, HeLa and SW480 cells were
transfected using different human SHOCA-2 (SH2D4A)-specific shRNA constructs (Origene).

**Bioinformatic analysis**

Sequences from SHOCA-1 (SH2D4B) and SHOCA-2 (SH2D4A) were retrieved by homology search with ENSEMBL: C. elegans SHOCA, F13B12.6; Danio rerio SHOCA-1, ENSDARG0000015144; Gallus gallus SHOCA-1, ENSGALG00000002370; Mus musculus SHOCA-1, ENSMUSG0000037833; H. sapiens SHOCA-1, ENSG00000178217; Danio rerio SHOCA-2, ENSDARG00000041378; Xenopus laevis SHOCA-2, ENSXETG0000004464; Gallus gallus SHOCA-2, ENSGALG0000010122; Mus musculus SHOCA-2, ENSMUSG00000053886; H. sapiens SHOCA-2, ENSG00000104811. The common SHOCA precursor (Sh2d4) sequence of Monosiga brevicollis (30425.AA) was identified in the KinBase database (http://www.kinase.com/kinbase/). Sequences were aligned with the TCOffee software (http://tcoffee.vital-it.ch). An unrooted evolutionary tree of SHOCA homologues was generated using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). The gene microarray analysis tool made available by Oncomine (Compendia Bioscience, Ann Arbor, MI, USA: http://www.oncomine.org) was used to determine SHOCA-2 mRNA expression in various cancers.

**Quantification of SHOCA-2 expression and STAT3 phosphorylation on tissue microarrays & whole-tissue sections**

The H score determines the percentage of positive tumor cells multiplied by their staining intensity. Given the size of the cohort of patients analyzed, three ranges of scores were
considered: >200, identifying strong expression or phosphorylation; 100-200, classifying moderate expression or phosphorylation; <100 characterizing a low or absent expression or phosphorylation.

**Statistical analyses**

The likelihood ratio test and binomial test for differences in proportions were used to compare H-scores between and among disease stages, respectively\(^5\). Survival time analysis was performed using three different methods: the non-parametric log-rank test, the non-parametric Gehan-Wilcoxon tests and the Cox regression analysis which was carried out after verification of the assumption of proportional hazards; survival time differences were represented using the Kaplan-Meier method\(^5\). Hazard ratios (HR) and 95\% confidence intervals (CI) were obtained to measure effect size with values <1.0 indicating more favorable survival time with higher protein expression. P-values were two-sided and considered statistically significant when p<0.05.

**Real-time RT-PCR**

Using TRI reagent (Molecular Research Center), total RNA was prepared from various tissues freshly isolated from 10 week-old Balb/c mice. Oligo(dT20) or random-N\(^6\)-primed cDNA was generated with M-MLV Reverse Transcriptase (Invitrogen-Gibco) employing standard protocols. For quantitative real-time PCR, the SYBR Green-based method was used (Sensimix, Quantace); primers are available upon request. PCR specificity was controlled by analyzing the melting curves and agarose gel electrophoresis of the PCR products. Amounts of specific mRNA were normalized to levels of 18S rRNA transcripts.
Mass spectrometry

TCA precipitated and acetone washed or Chlorophorum/Methanol treated protein pellets were reduced with TCEP (tris(2-carboxyethylphosphine), alkylated with Jodoacetamide and digested with Trypsin. The generated peptides were analyzed by NanoLC-MS/MS on a 4000Q Trap (MDS Sciex) as described⁷. The proteins were identified with Mascot (Matrix Science) searching the Uniprot database⁸. MRM-buddy, a software developed at the Friedrich Miescher Institut (A. di Cara, R. Sack and R. Portmann, unpublished results) was used to extract the MRM-relevant information for the quantification of selected proteins. MRM analysis was done as described⁷.

Western Blot and immunoprecipitation

Total protein (40µg as quantified by the Bradford method; Biorad) was separated on a 8%-12% SDS-PAGE and transferred to nitrocellulose membranes (Biorad). Membranes were blocked with 5% milk (Migros) in TBS-Tween and subsequently incubated overnight with primary antibody using concentrations that have either been previously established or used according to the manufacturer's recommendations. After a one hour incubation with a secondary antibody conjugated to horseradish peroxidase (Southern Biotech), the reaction was visualized by ECL (Pierce). For immunoprecipitation, cell lysates containing equal amounts of total protein were precleared for 1 hr with 40 µl of 50% (wt/vol) protein G Plus-Sepharose beads (Amersham) and incubated for 2-3 hrs at 4°C with primary or control (IgG) antibodies. The beads were added for 1 hr and then washed extensively with lysis buffer; bound proteins were fractionated on a 8%-12% SDS-PAGE, and analyzed by western blotting as detailed above. Immunoprecipitations were performed using RIPA buffer with the notable exception of analyses shown in Figures 3e and 4d.
which were carried out using the JS lysis buffer (50 mM HEPES at pH 7.5, 1% glycerol, 50 mM NaCl, 1% Triton X-100, 1.5 mM MgCl₂, 5 mM EGTA).

Confocal microscopy

HeLa cells grown plated on glass coverslips were transiently transfected to express the EGFP-PP1β fusion protein. Twenty-four hours later, cells were washed in PBS and fixed for 15 min using 4% paraformaldehyde at room temperature (RT). Following three washes with PBS, cells were permeabilized on ice for 15 min using cold 0.1% Triton X-100 (v/v) in PBS. All staining reactions were performed at RT. To prevent unspecific binding, the cells were exposed for 30 min to 5% normal goat serum diluted in PBS. Next, the cells were incubated for 1 hr with the primary antibody, then washed three times in PBS and finally exposed for 30 min to the secondary antibody, an Alexa 555-conjugated goat anti-mouse antibody (Invitrogen). Prior to mounting, nuclei were stained with DAPI (4',6-diamidino-2-phenylindole, diluted 1:10,000 in PBS from a 1 mg/ml stock) and washed twice with PBS. Coverslips were mounted with fluorescent mounting medium (Dako). Images were acquired using an LSM 510 Meta confocal microscope and LSM510 imaging software (Carl Zeiss).

Cell cycle analysis

HeLa cells stably transfected with shRNA plasmid specific for the knock-down of SHOCA-2 or with a control plasmid were serum-starved for 24 hrs prior to a 12 hrs stimulation with EGF (100 ng/ml). After trypsinization and centrifugation, 5 x 10⁵ cells were resuspended in 0.5 ml of cold hypotonic PI (Propidium Iodide) staining solution (50 µg/ml PI (Sigma), 0.1 mg/ml RNase A, 0.1% Triton X-100, 0.1% sodium citrate).
Following vigorous resuspension, the cells were incubated (15 min) on ice and then analyzed by flow cytometry (FACSCalibur, Becton Dickinson) using the FlowJo software (Tree Star, Oregon Corporation). G1, S and G2/M phases of cell cycle were defined using the mathematical Watson Pragmatic model.

**Cell proliferation**

SW480 cells (5 x 10^4 cells/well) were cultured in 96-well plates in the absence or presence of AG1478 (10 μM) or S31-201 (100μM) in 10% serum condition. After 48 hours of treatment, cell proliferation was analyzed using CellTiter 96 aqueous nonradioactive cell proliferation assay kit (Promega, Madison, WI). The cell proliferation index was calculated as a percentage of the absorbance in relation to the untreated control cells.

**Human tissues and DNA extraction**

The tissues from primary colorectal carcinomas and the macroscopically healthy-appearing neighboring mucosa (7cm proximal of tumor) were resected on the occasion of surgical cancer therapy. All the tissue samples were immediately submerged into RNALater® RNA stabilization reagent (Qiagen) and kept at room temperature for several hours to guarantee a complete penetration by the reagent. The vials were then stored at -80°C until the tissue was further processed. DNA from approximate 20 mg of tissue was extracted with the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer’s specifications.

**LOH and mutational analysis of the SH2D4A gene**
All canonical exons, the flanking intronic sequences as well as 5' and 3' UTR regions were subjected to PCR amplification (Kapa HiFi™ system; Kapa Biosystems, USA) and sequenced (BigDye terminator kit v1.1 on ABI 3130xl; both Applied Biosystems, USA) using the set of primers described in the supplementary table. Putative mutations were confirmed by bi-directional sequencing of a second independent PCR product. In Silico mutation impact prediction scores (SIFT: http://sift.jcvi.org and PolyPhen-2: http://genetics.bwh.harvard.edu/pph2) were calculated for all somatic point alterations found. Loss of heterozygosity was assessed using 3 flanking microsatellite markers (D8S549, D8S1715 and D8S258 labeled at 5' end with 6-Carboxyfluorescein (6-FAM) dye) as well as 6 single nucleotide polymorphisms refSNP rs2280444, rs877386, rs1574288, rs17128221, rs4921637, rs36092909) located within the SH2D4A gene itself. The relative proportions of alleles 1 and 2 in non-tumor and tumor samples corresponds to the ratio of the heights of the corresponding peaks. An LOH index was calculated by dividing the ratio of allele 2 to allele 1 in the non-tumor DNA by the corresponding ratio in the tumor DNA. LOH positivity was defined as an LOH index of <0.5 (reflecting a substantial loss of allele 1 in the tumor sample) or >1.5 (indicative of substantial loss of allele 2).

Identification and characterization of exon 4 skipping variant of the human SH2D4A gene by RT-PCR

First strand cDNA was synthesized from 500ng of total RNA using random-primed reverse transcription (Verso cDNA; Thermo Scientific). PCR amplification was next carried out on a cDNA template, products were separated on a 2% agarose gel and visualized with ethidium bromide. Individual bands were excised from the gel, eluted
(Quiquick gel extraction kit; Quiagen, Germany), reamplified and sequenced.

**Conditions and primers for sequence, LOH and cDNA analyses**

For the **SH2D4A** sequence analysis the following PCR conditions were used: initial denaturation of 2 min at 94°C followed by 35 cycles of 10 sec at 94°C, of 15 sec at 56°C and of 15 sec at 72°C and a final elongation step of 3 min at 72°C. When using the ABI Prism® TrueAllele® premix, the initial denaturation step was extended to 10 minutes. The following primers were used:

<table>
<thead>
<tr>
<th>Mutational Analysis</th>
<th>DNA sequence 5’→3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>SH2D4A_x1_For</td>
<td>CGATTGC GCCCGCCGCCAGTCA</td>
</tr>
<tr>
<td>SH2D4A_x1_Rev</td>
<td>CTCCCATGACACCCATGCAAC</td>
</tr>
<tr>
<td>SH2D4A_x2_For</td>
<td>GGAACCTTTTGGACAAGATAT</td>
</tr>
<tr>
<td>SH2D4A_x2_Rev</td>
<td>CTGCCACCTGCTTTGCTTG</td>
</tr>
<tr>
<td>SH2D4A_x3_For</td>
<td>ATTGCTACCTACAGATGTTCC</td>
</tr>
<tr>
<td>SH2D4A_x3_Rev</td>
<td>AAGTGA CTTTTGTGACCCCTC</td>
</tr>
<tr>
<td>SH2D4A_x4_For</td>
<td>ACGTCTTTACACACCTAGC</td>
</tr>
<tr>
<td>SH2D4A_x4_Rev</td>
<td>AAATCCGCACGTCTACAG</td>
</tr>
<tr>
<td>SH2D4A_x5_For</td>
<td>TCA CATGGCCGCTTGCTGT</td>
</tr>
<tr>
<td>SH2D4A_x5_Rev</td>
<td>ATGGCTAGGTATGCAAGCCC</td>
</tr>
<tr>
<td>SH2D4A_x6_For</td>
<td>AGCAAACGT GCTGATTGGCT</td>
</tr>
<tr>
<td>SH2D4A_x6_Rev</td>
<td>ATGACGCTGAGACCCTGGAAG</td>
</tr>
<tr>
<td>SH2D4A_x7_For</td>
<td>TAAGCAGACAACAAAACCTG</td>
</tr>
<tr>
<td>SH2D4A_x7_Rev</td>
<td>CTCATACAATTGCCCTTTAA</td>
</tr>
<tr>
<td>SH2D4A_x8_For</td>
<td>TATAGACCTTTTGAGGCC</td>
</tr>
<tr>
<td>SH2D4A_x8_Rev</td>
<td>TTCTTTGATGATAGCAACAG</td>
</tr>
<tr>
<td>SH2D4A_x9_For</td>
<td>TGCTCTATTTTGCTTATTTAGGC</td>
</tr>
<tr>
<td>SH2D4A_x9_Rev</td>
<td>CTACAAGGCAAAATTAAGTT</td>
</tr>
<tr>
<td>SH2D4A_x10_For</td>
<td>GGTCA CATG TGA TAAAGAAA</td>
</tr>
<tr>
<td>SH2D4A_x10_Rev</td>
<td>AAAACAAACTCAGCGATGT</td>
</tr>
</tbody>
</table>

**LOH Analysis**

<table>
<thead>
<tr>
<th>LOH Analysis</th>
<th>DNA sequence 5’→3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S1715_For</td>
<td>CAGGTGATGTCGCCCAAGG</td>
</tr>
<tr>
<td>D8S1715_Rev</td>
<td>CGAACATGAAATTGAAATCCAGTG</td>
</tr>
<tr>
<td>D8S258_For</td>
<td>CTGCCAGGAATCCAAGTACG</td>
</tr>
<tr>
<td>D8S258_Rev</td>
<td>TGAAGGACAGACCACG</td>
</tr>
<tr>
<td>D8S549_For</td>
<td>AAATGAACTCCTGATTGCCAAC</td>
</tr>
<tr>
<td>D8S549_Rev</td>
<td>TGAGAGCCAACTTATTTCTACC</td>
</tr>
<tr>
<td>rs36092909_For</td>
<td>ACAATGCCCAGCTGACATT</td>
</tr>
</tbody>
</table>
rs36092909_Rev | TACATCATTAAGGTCATAC
--- | ---
cDNA Analysis
SH2D4A_cx2.1_For | CAGACCAAAGAAAGAGAATGGCA
SH2D4A_cx2_Rev | ATCTGCCAACATTTTCGAA
SH2D4A_cx2.1_Rev | CTTTGATTTTTCGACAGATG

Multiplex-ligation dependent-probe-amplification (MLPA)

Chromosome 8 copy-number variation was investigated by multiplex ligation-dependent probe amplification assay (kit P014-1A, MRC Holland) according to the manufacturer’s protocols. Calculations were performed using GeneMarker software (SoftGenetics). Gene dosage with relative value of 1 is expected for two copies, whereas values of 0.5 and 1.5 indicate loss and gain of one copy, respectively. MLPA results indicative of a somatic deletion were independently confirmed in at least one additional experiment.

Methylation analysis

Genomic DNA from primary tumor samples and adjacent matched healthy appearing mucosa, respectively, was extracted using the QIAamp DNA Mini Kit (Qiagen), and bisulfite converted using the EZ DNA Methylation Kit™ (Zymo Research). Promoter associated CpG-island regions of SH2D4A were amplified by PCR and biotin labeled either directly using biotin-labelled reverse primers or during a second PCR step using a biotinylated universal primer annealing to a sequence introduced by the reverse primer used in the first PCR. PCR conditions and primer sequences are provided on supplementary table. Labelled PCR products were purified (QIAquick® PCR Purification Kit, Qiagen), bound to streptavidin-covered sepharose beads and analyzed by pyrosequencing using the PyroMark Q24 pyrosequencing system (Qiagen). Sequencing-assay design as well as quantification of methylation was performed with the PyroMark
Peripheral blood DNA treated with M.SssI methyltransferase (New England Biolabs) and subsequently bisulfite treated served as positive control.

### Conditions and primers for methylation analysis

<table>
<thead>
<tr>
<th>PCR Amplification of CpG islands</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1(^{st}) CpG island forward</td>
<td>5'-GTT TTA TGA ATT AAG AGA GGA GAA GAT ATT TTT ATT T-3'</td>
</tr>
<tr>
<td>1(^{st}) CpG island reverse</td>
<td>5'-GCC CCC GCC CGC CCC CAA CAC CTA CCA ATA C-3'</td>
</tr>
<tr>
<td>Biotinylated primer</td>
<td>5'-Bio-GCC CCC GCC CG-3'</td>
</tr>
<tr>
<td>2(^{nd}) CpG island forward</td>
<td>5'-AG GTT GTA TGG GTG TAT GGG AG-3'</td>
</tr>
<tr>
<td>2(^{nd}) CpG island biotinylated reverse</td>
<td>5'-Bio-CAA CTC AAA AAA CCT ACA AAC TAT AAC TCC TTC-3'</td>
</tr>
</tbody>
</table>

**PCR conditions:**

<table>
<thead>
<tr>
<th>1(^{st}) CpG island 1(^{st}) PCR</th>
<th>Enzyme activation: 95°C for 2min 40 cycles of  Denaturation: 95°C for 30 sec  Annealing: 61°C for 25 sec  Extension: 72°C for 30 sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(^{st}) CpG island 2(^{nd}) PCR</td>
<td>3.5mM MgCl(_2), 0.5(\mu)M primer  Enzyme activation: 95°C for 2min 20 cycles of  Denaturation: 95°C for 30 sec  Annealing: 61°C for 30 sec  Extension: 72°C for 30 sec</td>
</tr>
<tr>
<td>2(^{nd}) CpG island</td>
<td>2.5mM MgCl(_2), 0.5(\mu)M primer  Enzyme activation: 95°C for 4min 35 cycles of  Denaturation: 95°C for 30 sec  Annealing: 61°C for 30 sec  Extension: 72°C for 55 sec</td>
</tr>
</tbody>
</table>

**Pyrosequencing primers** |
| Primer 1   | 5'- GGA GGA GAG GAT AAG T -3' |
| Primer 2   | 5'- GTG TTG GAG AGT TTT TAG -3' |
| Primer 3   | 5'- GTT ATA AGA GTT TTG -3' |
| Primer 4   | 5'- GGT TAA GTG GAT GTG -3' |

| Primer 1   | 5'- TAT GGG TGT ATG GGA G -3' |
| Primer 2   | 5'- GGG TTT AGG TG -3' |
| Primer 3   | 5'- GTT GGT TAG GTG AG -3' |
| Primer 4   | 5'- AGG GAG AGG GTT TG -3' |
| Primer 5   | 5'- GGG GAG TGT AGG TTT -3' |
| Primer 6   | 5'- TAT GGA AAT TTT TTA GGA GG -3' |
| Primer 7   | 5'- TGG GTT TTG GGA G -3' |
| Primer 8   | 5'- GGG TTG GTT TAG ATT TAG -3' |

**Gel-Shift, Shift-Western Blot, and Western Blot**

The gel-shift method employed was adopted from Harrington et al. A 28 bp long, fluorescein-labeled dsDNA probe containing a Sp1 recognition site was incubated for 15 min at RT with SW48 nuclear extract in the presence of increasing concentrations of an unlabeled 28mer oligomer. The competitor had either a methylated or an unmethylated Sp1 recognition site. The mixture was then run in a 6% native polyacrylamide gel and fluorescence signals were subsequently detected using a biomolecular imager (Typhoon 9400, Amersham Biosciences). Fluorescence intensity of shifted bands was normalized against the sum of signals provided by both unbound oligomers and shifted bands. Experiments were performed in quadruplicates.

Bands from the polyacrylamide gel were transferred onto a nitrocellulose membrane using the wet blotting method and Sp1 was detected via chemiluminescence with anti-
human Sp1 primary antibody (polyclonal rabbit, Millipore) and a HRP-labeled anti-rabbit secondary antibody (GE healthcare).

Sequences of oligomers:
Positive control: Forward: 5′-ACG TTG CAG CCG GGG CCG TTC TGC A-3′,
Reverse: 5′-TGC AGA AGC CCC GCC CCG GCT GCA ACG T-3′
Methylated competitor: Forward: 5′-CCT CAC CCC CGC CTC CAC CCC TTC GCG G-3′,
Reverse: 5′-F-CCG CGA AGG GCT GGA GCC CCG GGT GAG G-3′
Unmethylated competitor: Forward: 5′-CCT CAC CCC CGC CTC CAC CCC TTC GCG G-3′, Reverse: 5′-F-CCG CGA AGG GCT GGA GCC CCG GGT GAG G-3′
Supplemental References


6.3 HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma.

My contribution to this work:
- Immunohistochemical analysis of whole tissue and microarray sections;
- Data analysis and statistical comparison;
- Manuscript writing;
HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma

Salvatore Piscuoglio, Inti Zlobec, Pierlorenzo Pallante, Romina Sepe, Francesco Esposito, Arthur Zimmermann, Ioannis Diamantis, Luigi Terracciano, Alfredo Fusco & Eva Karamitopoulou

Institute of Pathology, University of Basel, Basel, Switzerland, Istituto di Endocrinologia ed Oncologia Sperimentale del CNR c/o Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università degli Studi di Napoli 'Federico II', Naples, Italy, Institute of Pathology, University of Bern, Bern, and Department of Internal Medicine and Gastroenterology, Zofingen Hospital, AG, Switzerland

Date of submission 10 November 2010
Accepted for publication 9 February 2011


HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma

Aims: Pancreatic ductal adenocarcinoma follows a multistep model of progression through precursor lesions called pancreatic intraepithelial neoplasia (PanIN). The high mobility group A1 (HMGA1) and high mobility group A2 (HMGA2) proteins are architectural transcription factors that have been implicated in the pathogenesis and progression of malignant tumours, including pancreatic cancer. The aim of this study was to explore the role of HMGA1 and HMGA2 in pancreatic carcinogenesis.

Methods and results: HMGA1 and HMGA2 expression was examined in 210 ductal pancreatic adenocarcinomas from resection specimens, combined on a tissue microarray also including 40 examples of PanIN and 40 normal controls. The results were correlated with the clinicopathological parameters of the tumours and the outcome of the patients. The percentage of tumour cells showing HMGA1 and HMGA2 nuclear immunoreactivity correlated positively with increasing malignancy grade and lymph node metastasis. Moreover, HMGA1 and HMGA2 expression was significantly higher in invasive carcinomas than in PanINs. No, or very low, expression was found in normal pancreatic tissue.

Conclusions: Our results suggest that HMGA1 and HMGA2 are implicated in pancreatic carcinogenesis and may play a role in tumour progression towards a more malignant phenotype.

Keywords: HMGA1, HMGA2, immunoreactivity, pancreatic ductal adenocarcinoma, pancreatic intraepithelial neoplasia

Abbreviations: CI, confidence interval; HMGA, high mobility group A; NF-kB, nuclear factor-kB; PanIN, pancreatic intraepithelial neoplasia; ROC, receiver operating characteristic; SD, standard deviation; TMA, tissue microarray

Address for correspondence: A Fusco, Istituto di Endocrinologia ed Oncologia Sperimentale, del Consiglio Nazionale delle Ricerche, via Pansini 5, 80131 Naples, Italy. e-mail: alfusco@unina.it
L Terracciano, Institute of Pathology, University of Basel, Schönbeinstrasse 40, 4031 Basel. e-mail: lterracchano@uhbs.ch
© 2012 Blackwell Publishing Limited.
Introduction
Pancreatic ductal adenocarcinoma is a common cause of death from cancer, and has a dismal prognosis with currently no effective treatment. Despit important advances in our understanding of the molecular biology of the early stages of neoplastic development, late molecular events that lead to tumour progression are largely unknown. Clinicopathological parameters such as tumour size, lymph node metastases and evidence of blood vessel or lymphatic invasion have been proven to be reliable prognostic determinants in pancreatic cancer. The identification of reliable and reproducible biomarkers would enable better stratification of patients, and eventually provide a guide for individualized therapy. Pancreatic cancer follows a multistep model of progression through non-invasive precursor lesions. Pancreatic intraductal lesions have been classified into four groups of pancreatic intraepithelial neoplasias (PanINs): PanIN-1A, PanIN-1B, PanIN-2, and PanIN-3. PanIN-3 shows severe epithelial dysplasia, and is most likely to progress to invasive carcinoma. The high mobility group A (HMGA) genes encode a family of non-histone chromatin-binding proteins, named for their rapid electrophoretic mobility in polyacrylamide gels. HMGA1a and HMGA1b isoforms result from alternative splicing of HMGA1 mRNA, whereas HMGA2 is encoded by the related gene HMGA2. HMGA proteins bind the minor groove of AT-rich DNA sequences. Their DNA-binding domain is located in the N-terminal region of the protein, and contains three short basic repeats, the so-called AT-hooks. Once bound to DNA, the HMGA proteins alter chromatin structure and thereby regulate the transcriptional activity of several genes. HMGA proteins are normally expressed at high levels during embryonic development, and at very low levels in adult, differentiated tissues. HMGA proteins participate, as transcriptional regulators, in many cellular functions, including regulation of the cell cycle, cell differentiation, senescence, and neoplastic transformation. Both HMGA1 and HMGA2 have been reported to function as oncogenes and to be overexpressed in almost all human malignancies so far analysed, including ductal pancreatic adenocarcinoma. Moreover, HMGA protein overexpression has been regarded as a poor prognostic feature, as it has often been found to correlate with the presence of metastasis and with reduced survival.

The objective of the present study was to investigate the role of HMGA1 and HMGA2 expression in pancreatic carcinogenesis and to evaluate its prognostic significance. Using immunohistochemistry we analysed expression in different stages of pancreatic carcinogenesis, including invasive adenocarcinomas, PanINs, and normal pancreatic tissue, in a tissue microarray (TMA) combining 210 ductal adenocarcinomas of the pancreas from resection specimens, 40 cases of PanIN-3, and 40 normal controls.

Materials and methods
Patients and specimens
Formalin-fixed and paraffin-embedded tumours and control specimens were retrieved from the archives of the Institute of Pathology, University of Bern. All tumours and controls were reviewed by an experienced pathologist (E.K.). Histological subtypes other than ductal carcinoma were excluded. Tumours were re-staged according to the American Joint Committee on Cancer Staging Manual (seventh edition). Representative tumour areas were selected for the construction of the TMA. The TMA consisted of 210 surgically-resected ductal adenocarcinomas of the pancreas, and included 40 examples of PanIN-3 and 40 normal controls (normal pancreatic tissue and PanINs were selected from areas distant from the carcinomas). The 210 patients comprised 110 males and 100 females, with a mean age of 66.5 years (range: 20–92 years). The study was approved by the ethics committee of the University of Bern.

Assessment of behaviour
Medical charts were available for 77 of the 210 patients. Of these 77 patients, 60 (78%) died from the disease, and 7 (9%) were alive with recurrent/metastatic disease. The other 10 patients (13%) were alive without disease. The median follow-up time was 16 months. The clinicopathological features of these cases with survival information are given in Table 1.

Construction of the TMA
One core tissue biopsy with a diameter of 0.6 mm was taken from a representative region of individual paraffin-embedded pancreatic carcinomas (donor blocks), and placed into a new recipient paraffin block with a semi-automated tissue-arraying device. The presence of tumour tissue on the TMA was verified on a haematoxylin and eosin-stained slide. Two to three tissue cores of each tumour were available for biomarker analysis. Five-micrometre sections were cut with an adhesive-coated slide system (Instrumedics, Hackensack, NJ, 2012 Blackwell Publishing Ltd, Histopathology, 60, 397–404.
and then with the primary antibody at room temperature. Optimal staining was achieved after pretreatment in a microwave oven (98°C for 30 min, pH 6, dilution 1:250). Subsequently, sections were incubated with peroxidase-labelled secondary antibody (Dako Cytomation) for 30 min at room temperature. 3,3′-Diaminobenzidine was used as chromogen. Sections were then counterstained with Gill’s haematoxylin. As a positive control, a TMA with various normal tissue samples was stained in parallel.

The antibodies used for HMGA1 immunostaining were raised against the synthetic peptide SSSKQQPLASKQ, which is specific for HMGA1. They were affinity purified against the synthetic peptide.9 For HMGA2 immunohistochemistry, antibodies raised against a synthetic peptide located in the N-terminal region were used.22

The specificity of immunolabelling was validated by the absence of tumour staining when using antibodies preincubated with the peptide against which the antibodies were raised (data not shown). Similarly, no positivity was observed when tumour samples were incubated with a preimmune serum (data not shown).

**IMMUNOHISTOCHEMICAL EVALUATION**

Nuclear HMGA1 and HMGA2 staining was scored by two independent observers (S.P. and L.T.) blinded for clinical parameters. Slides were screened semiquantitatively for the percentage of positive cells and the intensity of the signal. At least 100 cells were counted for each punch. The percentage of positive cells per number of cells counted was scored in 10 groups from 0 (0–9%) to 9 (91–100%). The intensity of the signal was graded semiquantitatively into four groups from 0 (no positivity) to 3 (strong positivity). A case was considered to be positive if belonging at least to group 1 for the percentage (i.e. ≥10%), irrespective of intensity. In PanINs and normal controls, the epithelial cells of ductal structures were evaluated.

**STATISTICAL METHODS**

The selection of clinically important cut-off scores was based on receiver operating characteristic (ROC) curve analysis.23,24 At each percentage score, the sensitivity and specificity for each outcome under study were plotted, generating an ROC curve. The score having the closest distance to the point with both maximum sensitivity and specificity, i.e. point (0.0, 1.0) on the curve, was selected as the cut-off score leading to the greatest number of tumours that were correctly classified as having or not having the outcome. In order to

### Table 1. Clinicopathological characteristics of cases with survival information (N = 77)

<table>
<thead>
<tr>
<th>Clinicopathological features</th>
<th>Frequency, N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnosis</strong></td>
<td></td>
</tr>
<tr>
<td>Ductal carcinoma</td>
<td>77 (100.0)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>33 (42.9)</td>
</tr>
<tr>
<td>Male</td>
<td>44 (57.1)</td>
</tr>
<tr>
<td><strong>Tumour grade</strong></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>16 (20.8)</td>
</tr>
<tr>
<td>G2</td>
<td>42 (54.6)</td>
</tr>
<tr>
<td>G3</td>
<td>19 (24.7)</td>
</tr>
<tr>
<td><strong>pT stage</strong></td>
<td></td>
</tr>
<tr>
<td>pT1</td>
<td>3 (4.1)</td>
</tr>
<tr>
<td>pT2</td>
<td>12 (16.2)</td>
</tr>
<tr>
<td>pT3</td>
<td>52 (70.3)</td>
</tr>
<tr>
<td>pT4</td>
<td>7 (9.5)</td>
</tr>
<tr>
<td><strong>pN stage</strong></td>
<td></td>
</tr>
<tr>
<td>pN0</td>
<td>27 (38.0)</td>
</tr>
<tr>
<td>pN1</td>
<td>44 (62.0)</td>
</tr>
<tr>
<td><strong>Metastasis</strong></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>72 (93.5)</td>
</tr>
<tr>
<td>Present</td>
<td>5 (6.5)</td>
</tr>
<tr>
<td><strong>Tumour diameter (mm), mean ± SD</strong></td>
<td>31.4 ± 14.0</td>
</tr>
<tr>
<td><strong>Survival time (months), median (range)</strong></td>
<td>12.0 (0.5–48.0)</td>
</tr>
</tbody>
</table>

SD, Standard deviation.

USA) and examined by immunohistochemistry. The number of samples differed slightly between the individual markers, because of variability in the number of interpretable specimens on TMA sections.

**IMMUNOHISTOCHEMISTRY**

Freshly cut sections of TMA blocks were used for immunohistochemical staining with anti-HMGA1 and anti-HMGA2 antibodies. Briefly, punches were de-waxed and rehydrated in distilled water. Endogenous peroxidase activity was blocked with 0.5% H2O2. The sections were incubated with 10% normal goat serum (Dako Cytomation, Carpinteria, CA, USA) for 20 min, and then with the primary antibody at room temperature. Optimal staining was achieved after pretreatment in a microwave oven (98°C for 30 min, pH 6, dilution 1:250). Subsequently, sections were incubated with peroxidase-labelled secondary antibody (Dako Cytomation) for 30 min at room temperature. 3,3′-Diaminobenzidine was used as chromogen. Sections were then counterstained with Gill’s haematoxylin. As a positive control, a TMA with various normal tissue samples was stained in parallel.

The antibodies used for HMGA1 immunostaining were raised against the synthetic peptide SSSKQQPLASKQ, which is specific for HMGA1. They were affinity purified against the synthetic peptide.9 For HMGA2 immunohistochemistry, antibodies raised against a synthetic peptide located in the N-terminal region were used.22

The specificity of immunolabelling was validated by the absence of tumour staining when using antibodies preincubated with the peptide against which the antibodies were raised (data not shown). Similarly, no positivity was observed when tumour samples were incubated with a preimmune serum (data not shown).

**IMMUNOHISTOCHEMICAL EVALUATION**

Nuclear HMGA1 and HMGA2 staining was scored by two independent observers (S.P. and L.T.) blinded for clinical parameters. Slides were screened semiquantitatively for the percentage of positive cells and the intensity of the signal. At least 100 cells were counted for each punch. The percentage of positive cells per number of cells counted was scored in 10 groups from 0 (0–9%) to 9 (91–100%). The intensity of the signal was graded semiquantitatively into four groups from 0 (no positivity) to 3 (strong positivity). A case was considered to be positive if belonging at least to group 1 for the percentage (i.e. ≥10%), irrespective of intensity. In PanINs and normal controls, the epithelial cells of ductal structures were evaluated.

**STATISTICAL METHODS**

The selection of clinically important cut-off scores was based on receiver operating characteristic (ROC) curve analysis.23,24 At each percentage score, the sensitivity and specificity for each outcome under study were plotted, generating an ROC curve. The score having the closest distance to the point with both maximum sensitivity and specificity, i.e. point (0.0, 1.0) on the curve, was selected as the cut-off score leading to the greatest number of tumours that were correctly classified as having or not having the outcome. In order to
enable the use of ROC curve analysis, the following clinicopathological features were dichotomized: T stage (early, T1 + T2; late, T3 + T4), N stage (N0, no lymph node involvement; N1, any lymph node involvement), tumour grade (low, G1 + G2; high, G3), and survival (death from pancreatic carcinoma or alive).

Chi-square tests were used to study the relationship between HMGA1 and HMGA2 expression and histological subgroups. Differences in HMGA1 and HMGA2 expression between normal tissue, PanIN and carcinoma were investigated with the non-parametric Wilcoxon rank sum test. Univariate survival analysis was carried out with the Kaplan–Meier log-rank test, and multivariate analysis with Cox proportional hazards regression. Hazard ratios and 95% confidence intervals (CIs) were used to determine the effect of each variable on survival time. In addition, logistic regression was performed in univariate and multivariate settings to determine the associations of protein expression and its independent effect on binary outcomes. The odds ratios and 95% CIs were evaluated. A Bonferroni correction for multiple comparisons was performed. A $P$-value $\leq 0.01$ (two-sided) was required for the association to be statistically significant. All analyses were carried out with SAS (V9; SAS Institute, Cary, NC, USA).

**Results**

**Analysis of HMGA1 and HMGA2 Expression by Immunohistochemistry**

Table 2 shows the differences in protein expression between normal pancreas, PanIN, and cancer, and Table 3 the correlation of HMGA1 and HMGA2 expression with pT stage, pN stage, and tumour grade. Survival related to protein marker expression is analysed in Table 4. Some representative images of immunohistochemical staining are illustrated in Figures 1 and 2.

**Pancreatic Carcinomas versus Normal Controls**

Pancreatic carcinoma cases generally displayed strong nuclear HMGA1 and HMGA2 immunoreactivity, whereas absent or very low immunoreactivity was observed in normal pancreatic tissue. The mean $\pm$ standard deviation (SD) for the percentage of cells showing HMGA1 and HMGA2 protein expression were found to be $0 \pm 0$ and $0.2 \pm 0.9$, respectively, in normal tissue, as compared with $26.6 \pm 30.5$ and $16.3 \pm 28.4$, respectively, in carcinomas ($P < 0.001$; Table 2).

**Pancreatic Carcinoma versus PanIN**

Percentages of cells showing HMGA1 and HMGA2 protein expression were significantly higher in ductal pancreatic adenocarcinomas ($\text{mean} \pm \text{SD}: 26.6 \pm 30.5$ and $16.3 \pm 28.4$, respectively) than in PanIN cases ($11.1 \pm 15.0$ and $2.7 \pm 13.5$, respectively) ($P < 0.001$; Table 2).

**PanIN versus Normal Controls**

Means for the percentages of cells showing HMGA1 and HMGA2 protein expression were significantly higher in PanIN cases ($11.1 \pm 15.0$ and $2.7 \pm 13.5$, respectively) than in normal tissue ($0 \pm 0$ and $0.2 \pm 0.9$, respectively) ($P < 0.001$; Table 2).

### Table 2. Differences in marker expression between normal pancreas, pancreatic intraepithelial neoplasia (PanIN) and cancer

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>PanIN</th>
<th>Cancer</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HMGA1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>31</td>
<td>31</td>
<td>183</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Mean $\pm$ SD</td>
<td>0 $\pm$ 0</td>
<td>11.1 $\pm$ 15.0</td>
<td>26.6 $\pm$ 30.5</td>
<td></td>
</tr>
<tr>
<td>Median (min–max)</td>
<td>0 (0–0)</td>
<td>5.0 (0–50)</td>
<td>10.0 (0–100)</td>
<td></td>
</tr>
<tr>
<td><strong>HMGA2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>29</td>
<td>37</td>
<td>191</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Mean $\pm$ SD</td>
<td>0.2 $\pm$ 0.9</td>
<td>2.7 $\pm$ 13.5</td>
<td>16.3 $\pm$ 28.4</td>
<td></td>
</tr>
<tr>
<td>Median (min–max)</td>
<td>0 (0–5)</td>
<td>0 (0–80)</td>
<td>0 (0–100)</td>
<td></td>
</tr>
</tbody>
</table>

HMGA, High mobility group A; SD, standard deviation. Expression is given as percentage of immunoreactive cells. Wilcoxon rank sum test.
Table 3. Protein marker expression related to stage (pT, pN), and tumour grade

<table>
<thead>
<tr>
<th></th>
<th>HMGA1</th>
<th>HMGA2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>pT stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pT1–2</td>
<td>25.7 ± 28.7; 15.0</td>
<td>16.0 ± 28.4; 0.0</td>
</tr>
<tr>
<td>pT3–4</td>
<td>27.1 ± 31.3; 10.0</td>
<td>16.9 ± 28.9; 0.0</td>
</tr>
<tr>
<td><strong>P-value</strong></td>
<td>0.83</td>
<td>0.952</td>
</tr>
<tr>
<td><strong>pN stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pN0</td>
<td>18.3 ± 22.8; 5.0</td>
<td>8.4 ± 19.0; 0.0</td>
</tr>
<tr>
<td>pN1</td>
<td>32.7 ± 33.6; 20.0</td>
<td>21.5 ± 32.4; 0.0</td>
</tr>
<tr>
<td><strong>P-value</strong></td>
<td>0.012</td>
<td>0.039</td>
</tr>
<tr>
<td><strong>Tumour grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1–2</td>
<td>22.2 ± 27.3; 10.0</td>
<td>12.0 ± 23.3; 0.0</td>
</tr>
<tr>
<td>G3</td>
<td>37.3 ± 35.1; 30.0</td>
<td>26.8 ± 36.0; 5.0</td>
</tr>
<tr>
<td><strong>P-value</strong></td>
<td>0.009</td>
<td>0.008</td>
</tr>
</tbody>
</table>

HMGA, High mobility group A.
Expression is given as the percentage of immunoreactive tumour cells. Mean ± standard deviation; median values. Wilcoxon rank sum test.

Table 4. Survival analysis related to protein markers using cut-off scores [median values, namely 10% for HMGA1 (nuclear), and 0% for HMGA2 (nuclear)]*

<table>
<thead>
<tr>
<th></th>
<th>Total no. of patients</th>
<th>No. of deaths</th>
<th>Median survival time (months) (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HMGA1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>40</td>
<td>30</td>
<td>14 (10–17)</td>
<td>0.816</td>
</tr>
<tr>
<td>Positive</td>
<td>24</td>
<td>20</td>
<td>12.5 (10–22)</td>
<td></td>
</tr>
<tr>
<td><strong>HMGA2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>37</td>
<td>27</td>
<td>15 (10–24)</td>
<td>0.196</td>
</tr>
<tr>
<td>Positive</td>
<td>26</td>
<td>21</td>
<td>14 (12–18)</td>
<td></td>
</tr>
</tbody>
</table>

CI, Confidence interval; HMGA, high mobility group A. Log-rank test.
*Similar results were obtained when expression was analysed as a continuous variable by Cox regression analysis.

Figure 1. Examples of nuclear immunohistochemical detection of high mobility group A1 (HMGA1). Absent expression in normal pancreatic tissue (A) and moderate expression in pancreatic intra-epithelial neoplasia (PanIN) (B), compared with strong, diffuse expression in pancreatic carcinoma (C).

Protein expression and tumour grading

Extent of nuclear HMGA1 and HMGA2 expression showed a positive correlation with higher tumour grade. Cells within poorly differentiated tumours (grade 3) more frequently expressed HMGA1 and HMGA2 (mean percentage expression ± SD: 37.3 ± 35.1 and 26.8 ± 36.0, respectively) than those in better-differentiated (grade 1 and 2) tumours (22.2 ± 27.3 and 12.0 ± 23.3, respectively) \((P = 0.009\) and \(P = 0.008\), respectively; Table 3).

Protein expression and TNM classification of the tumours

HMGA1 and HMGA2 expression showed a significant association with the pN stage of the tumours. Mean ± SD percentage expression levels for HMGA1 and HMGA2 were 18.3 ± 22.8 and 8.4 ± 19.0, respectively, for nodal negative (pN0) carcinomas, as compared with 32.7 ± 33.6 and 21.5 ± 32.4, respectively, for nodal positive (pN1) carcinomas \((P = 0.012\) and \(P = 0.039\), respectively; Table 3). No association was noted between protein expression and pT stage of the tumours \((P = 0.83\) and \(P = 0.952\), respectively; Table 3).

Prognostic significance

Median survival times were 12.5 and 14 months for HMGA1-positive and HMGA2-positive tumours, respectively, as compared with 14 and 15 months for HMGA1-negative and HMGA2-negative tumours, respectively. These differences were not statistically significant \((P = 0.816\) and \(P = 0.196\), respectively; Table 4).

Discussion

Although most patients with pancreatic cancer present with advanced disease, the molecular events involved in tumour progression, invasion and metastasis are poorly understood.

In the present study, we investigated the immunohistochemical expression of HMGA1 and HMGA2 in 210 cases of ductal adenocarcinoma of the pancreas combined on a TMA including in addition 40 examples of PanIN-3 and 40 normal controls.

A major finding was the increasing mean protein expression of HMGA1 and HMGA2 between normal pancreatic tissue, PanIN cases, and invasive adenocarcinoma (Table 2). Mean HMGA1 and HMGA2 expression appeared to progressively increase through the
transition from normal tissue to pancreatic cancer. Moreover, HMGA1 and HMGA2 expression showed a positive correlation with malignancy grade and neoplastic progression, becoming higher with the dedifferentiation of the neoplasms and with the presence of lymph node metastasis. Therefore, our data suggest that HMGA1 and HMGA2 expression correlates with a more aggressive phenotype in pancreatic adenocarcinoma. These findings are in keeping with the recent studies of Hristov et al., who also noted an association between HMGA1 and HMGA2 expression and a more malignant phenotype in pancreatic cancer. HMGA1 was found to correlate with advanced tumour grade and decreased survival of their patients, whereas HMGA2 correlated with increasing tumour grade and lymph node metastasis. However, in our study, both HMGA1 and HMGA2 showed a positive correlation with lymph node metastasis. Although we found that patients with HMGA1 and HMGA2 negative tumours tended to survive longer, the association with patient outcome was not statistically significant. This may be because of the short survival time of most patients with pancreatic cancer. In addition, Hristov et al. reported that HMGA1 expression correlated with a more advanced PanIN grade, whereas possible differences in protein expression between PanIN and adenocarcinoma were not discussed. In our study, only PanIN-3 lesions were included. Moreover, we demonstrated that mean protein expression was significantly higher in pancreatic ductal adenocarcinomas than in PanIN lesions.

A number of other groups have reported involvement of HMGA genes and proteins in the pathogenesis of pancreatic cancer. In one study, HMGA1 was found to be overexpressed in a small number of pancreatic adenocarcinomas and metastatic lesions, but without association with tumour grade. More recently, Liau et al. reported HMGA1 protein positivity in a high proportion of ductal pancreatic adenocarcinomas, also without correlation with tumour differentiation. However, in this study, in contrast to ours, only staining intensity was analysed, and the number of carcinomas was smaller. HMGA1 expression was also found to be increased in other pancreatic tumours, such as intraductal papillary mucinous neoplasms. Studies in pancreatic ductal adenocarcinoma cell lines have shown that HMGA1 knockdown decreases cellular invasion, anchorage-independent cell growth, and resistance to chemotherapeutic agents. In addition, high-level expression of HMGA1 has been reported in almost all neoplastic tissues, including colon, breast, lung, ovarian, uterine, prostatic, gastric and head and neck carcinomas. HMGA2 has also been implicated in the development and progression of human malignancies, including lung adenocarcinomas, breast cancer, and squamous cell carcinomas of the oral cavity. Additionally, HMGA2 has been reported to play a role in the epithelial–mesenchymal transition that takes place during invasion and metastasis. A previous study found increased HMGA2 expression by reverse transcription polymerase chain reaction in ductal pancreatic adenocarcinomas, and HMGA2 expression by immunohistochemistry. However, associations with grade and outcome were not included in the analysis, and HMGA2 expression was also found in pancreatic islet cells and, focally, in non-neoplastic ductal epithelial cells. In our study, focal HMGA2 immunoreactivity was also observed in a very small number of non-malignant ductal epithelial cells.

Regarding the processes underlying the involvement of HMGA genes in neoplastic transformation, it has been hypothesized that this probably occurs through oncofetal transcriptional mechanisms that have not yet been characterized. It has been suggested that the elevated expression of HMGA1 in tumour cells requires a complex cooperation between SP1 family members and AP1 factors, induced by the activation of Ras GTPase signalling. Moreover, the main function of the HMGA proteins, the regulation of gene transcription, is probably based on the ability of HMGA proteins to down-regulate or up-regulate the expression of genes that have a crucial role in the control of cell proliferation and invasion. In particular, emerging evidence suggests that HMGA1 modulates gene expression, including pathways involved in inflammation, proliferation, transformation, metastatic progression, angiogenesis, and DNA repair. Most transcriptional targets include regulatory elements of nuclear factor-kB (NF-kB), a mediator of inflammatory pathways, suggesting that HMGA1 and NF-kB may cooperate to induce inflammatory signals and drive transformation.

One of the advantages of this study is the use of TMAs, which have provided us with an efficient and cost-effective way of testing a large number of tumour specimens. Concerns could be raised about the TMA technique with regard to the possible limitations in sampling large, heterogeneous tumours. However, previous studies have shown comparable results between whole tissue sections and TMA cores, and have been able to reproduce numerous clinicopathological associations previously found with whole tissue sections.

In conclusion, we found increasing mean HMGA1 and HMGA2 expression during neoplastic progression in ductal pancreatic adenocarcinoma, accompanied by
a positive correlation of protein expression with both increasing malignancy grade and the presence of lymph node metastasis. Our results support the idea that HMGA1 and HMGA2 may play a significant role in the late stages of pancreatic carcinogenesis and in the progression towards a more aggressive tumour phenotype.

References

6.3.1 HMGA1 over-expression represents a poor prognostic index in human breast carcinoma.

My contribution to this work:
- Immunohistochemical analysis of whole tissue and microarray sections;
- Data analysis and statistical comparison;
- Manuscript writing;
HMGA1 over-expression represents a poor prognostic index in human breast carcinomas

Pierlorenzo Pallante¹, Salvatore Piscuoglio²,³*, Inti Zlobec², Luigi Terracciano² and Alfredo Fusco¹

¹Istituto di Endocrinologia ed Oncologia Sperimentale del CNR c/o Dipartimento di Biologia e Patologia Cellulare e Molecolare, Facoltà di Medicina e Chirurgia di Napoli, Università degli Studi di Napoli “Federico II”, via Pansini 5, 80131 Naples, Italy.

²Institute of Pathology, Molecular Pathology Division, University of Basel, Schonbeinstrasse 40, 4003 Basel, Switzerland.

³Research Group Human Genetics, Department of Biomedicine, University of Basel, Mattenstrasse 28, 4058 Basel, Switzerland

*equally contributed

Running title: HMGA1 in breast carcinomas.

Keywords: HMGA1, breast ductal carcinomas, immunohistochemistry.

Alfredo Fusco,
Istituto di Endocrinologia ed Oncologia Sperimentale del CNR,
via Pansini 5, 80131 Napoli, Italy.
Tel: 39-081-7463602 or 7463749;
Fax: 39-081-2296674.
E-mail: afusco@napoli.com or alfusco@unina.it

and

Luigi Terracciano,
Department of Pathology, Molecular Pathology Division,
University of Basel, Schonbeinstrasse 40, 4003 Basel, Switzerland.
Tel: 41-61-26522525 or 2652849;
Fax: 41-61-2653194.
E-mail: lterracciano@uhbs.ch
ABSTRACT

Aim: Breast cancer represents the second leading cause of cancer mortality among women and accounts for more than 40,000 deaths annually. HMGA1 expression has been implicated in the pathogenesis and progression of human malignant tumours, including breast carcinomas. The aim of this study was to evaluate HMGA detection as a possible prognostic index in breast carcinoma by analyzing a large number of breast carcinoma samples.

Methods: HMGA1 expression has been analyzed by immunohistochemistry in a large series of breast carcinoma resections (n=1202) combined on a Tissue Microarray (TMA) mainly including the ductal carcinoma variant. Then, the results were correlated with clinic-pathological parameters and outcome of the patients.

Results: HMGA1 over-expression was found in the large majority of breast carcinoma samples, and its over-expression positively correlates with Her2/neu amplification and progesterone receptor status, while a negative correlation was found with estrogen receptor status. Conversely, no HMGA1 expression was found in normal breast tissues.

Conclusions: The data reported here indicate that the level of HMGA1 expression is related to an unfavourable breast cancer phenotype and poor prognosis, as supported by its strong association with the Her2/neu, PR and ER status that could explain, at least in part, the different behaviour of the human breast carcinoma over-expressing HMGA1.
INTRODUCTION

Breast cancer represents the second leading cause of mortality caused by cancer in women (Ellis et al. 2003). Neoplastic breast diseases comprise benign form, like fibroadenoma, and very aggressive forms, like undifferentiated breast carcinoma. It has been reported that a large series of molecules and patterns (growth factors and their receptors, signal transduction molecules, cell cycle regulators) are altered and deregulated in sporadic breast cancers (Vogelstein and Kinzler 1994).

Nowadays, a series of genetic markers are evaluated to assess the prognosis of breast cancer patients: the BRCA mutational status of patients, the expression of estrogen receptor (ER), progesterone receptor (PR), and the Her2/neu receptor (Her2) (Deroo et al. 2006, Gao et al. 2002, Hynes et al. 1994, Miecznikowski et al. 2010). More recently, TGF-beta has also been considered as a potential prognostic marker for breast cancer patients (Koumoundourou et al. 2007). Moreover, the genetic status has a critical role in assigning the treatment. Indeed, tamoxifen (anti-estrogen agent) and trastuzumab (monoclonal antibody) are the elected chemotherapeutic agents for treating estrogen receptor (ER)–positive breast tumors and human Her2–over-expressing tumors, respectively (Carter et al. 1992, Arteaga et al. 2003). However, these markers are insufficient to predict the prognosis and to indicete the appropriate therapy, and many patients remain over- or under-treated (Cianfrocca and Gradishar, 2009).

HMGA1 protein belongs to the high-mobility group A (HMGA) family that consists of three members: HMGA1a, HMGA1b and HMGA2. Two distinct genes, HMGA1a and HMGA1b generates these three proteins by alternative splicing (Johnson et al. 1989). These proteins are able to bind AT-rich DNA sequences, but do not have transcriptional activity per se. However they can alter chromatin structure, therefore modulating the transcriptional activation of genes (Thanos and Maniatis 1992, Grosschedl et al. 1994).

HMGA over-expression is a feature of malignant tumours. Both HMGA genes are widely expressed during embryogenesis and in neoplastic tissues (including pancreas, thyroid, colon, breast, lung, ovary, uterine cervix, prostate, gastric carcinomas, squamous carcinomas of the oral cavity, head and neck tumours), whereas their expression is absent or very low in adult tissues. Their over-
expression represents a poor prognostic index and often correlates with metastases and reduced survival (Fusco and Fedele, 2007). Their oncogenic role has been extensively reported (Wood et al. 2000, Reeves et al. 2001, Berlingieri et al. 2002).

Previous studies of our group performed on a limited number of breast carcinoma samples, of which the clinic-pathological data were available only in a small number of cases, did not show any association between HMGA1 expression and histological grading. Conversely, we found that HMGA1 expression tended to be associated with c-erbB-2 expression (Spearman rho: 0.36; p=0.065), but not with the expression of the receptors for estrogens and progesterone. These results appeared in contrast with previous results, showing that enforced expression of HMGA1 in breast carcinoma cells induced the ability to form primary and metastatic tumours in athymic mice, and that HMGA1 is able to bind to BRCA1 promoter down-regulating its expression (Baldassarre et al. 2003).

The aim of the present work was, therefore, to analyse HMGA1 expression in a very large number of breast carcinoma tissues (n=1024), all of them provided with the most important clinicopathological parameters of patients, such as tumour size, lymph node status, endocrine receptors and Her2 status. Here, we confirm the HMGA1 over-expression in human breast carcinomas, with respect to the normal breast tissues. Moreover, we found that a strong HMGA1 over-expression identifies a subset of breast carcinomas characterized by Her2/neu amplification and progesterone receptor (PR) expression, but lacking estrogen receptor (ER) expression, a signature correlated with a patient poor prognosis.
MATERIALS AND METHODS

Human breast tissue samples

Neoplastic human breast tissue samples and normal controls were collected at the Department of Pathology, University of Basel Switzerland. Specimen obtained from surgical resections were formalin-fixed and paraffin embedded and stored in the archive of the institute. Several expert pathologists (SP, LT and LMT) proceeded to examine and classify tumours according to the American Joint Committee on Cancer (AJCC) staging manual (sixth edition). Tumour areas were accurately selected to be representative of the tumor specimen in the construction of the TMA. The TMA consisted of n=1388 “punches”, n=1202 of which (%) were evaluable. Diverse tumoral histotype, as well as clinicopathological characteristic of breast carcinoma samples are precisely indicated in Table 1.

Construction of tissue microarray

Construction of TMA was reported elsewhere (Torhorst et al., 2001). Briefly, representative regions of paraffin-embedded donor blocks corresponding to a single tumor specimen were used to take core biopsies with 0.6 mm diameter. Subsequently, semi-automated apparatus for tissue arraying (Beecher Instruments, Silver Spring, MD) was used to transfer core biopsies into a new recipient. Sections of 5 micrometer thick were cut using an adhesive-coated slide system (Instrumedics Inc., Hackensack, NJ) and analyzed by immunohistochemistry, after verification of TMA by H&E staining.

Immunohistochemistry and immunohistochemical evaluation

TMA sections were used for immunohistochemical analysis of HMGA1 protein expression by using an antibody raised against the N-terminal region of the HMGA1 protein as described elsewhere (Chiappetta et al. 2004). Staining procedures are performed as reported elsewhere (Piscuoglio et al. 2012). Negative controls, to confirm the specificity of the reaction, were performed by omitting the first antibody and by pre-incubating of the first antibody with molar excess of the HMGA1 synthetic peptide. Standard indirect immunoperoxidase procedure (ABC Elite, Vector Laboratories, Burlingame, CA) was used to develop signals. HMGA1 staining was scored by three expert pathologists (SP, LT and LMT) blinded for the clinic-
pathological parameters. A semi-quantitative methodology was used to perform the screening for the percentage of positive cells and for the signal intensity. At least 100 cells were counted in each punch.

**Statistical analysis**

Statistical correlations between variables were tested using a T student test (paired and unpaired), where appropriate. All tests were two-sided. A p-value <0.05 was considered statistically significant. Analysis was performed using SAS V9.1 (SAS Institute, Cary, NC, USA).

**Ethics**

All the analyses of this study were performed according to the ethical standards required by the local ethic committee of Dipartimento di Biologia e Patologia Cellulare e Molecolare, Universita' degli Studi di Napoli “Federico II”, IEOS-CNR, Napoli, Italy, and Department of Pathology, University of Basel, Basel, Switzerland.
RESULTS

Analysis of HMGA1 expression by immunohistochemistry.

A tissue micro array (TMA) comprising n=1338 cases of human breast carcinomas was analyzed by immunohistochemistry for HMGA1 protein expression using specific antibodies raised against the N-terminal region of the HMGA1 protein as described elsewhere (Chiappetta et al. 2004). TMA consisted mainly of ductal carcinoma tissues (n=963), and also including several normal breast tissue samples as positive controls. The whole set of clinic-pathological features of the patients are reported in Table 1. A total of n=1202 samples were informative for determination of the HMGA1 expression pattern that resulted always nuclear.

We found that HMGA1 staining was negative in all the normal cases analyzed, while HMGA1 expression was positive in all the breast carcinoma tissues analyzed (Figure 1). As shown in Table 1, no association was found between HMGA1 positivity and the ductal or other histotypes (p=0.883) of breast carcinomas. Indeed, HMGA1 staining, expressed as percentage (mean % ± sd, median %) of HMGA1-positive cells, was similar among ductal breast carcinomas with different histologic grading (Table 1). No significant association was also found with the pT stage of the tumours (p=0.402, Table 1). As well, no association was observed between HMGA1 protein expression and BRE grade of patients (p=0.2, Table 1).

Conversely, an inverse correlation was found between HMGA1 over-expression and the lymph node status of patients (p=0.293, Table 1). In fact, breast carcinomas not showing lymph-node colonization (pN0) expressed higher HMGA1 levels than their counterpart >pN0. However, this result failed to reach the statistical significance (Table 1), but if we consider only pN0, pN1 and pN2, this relationship is statistically significant (Table 1).

Therefore, these findings indicate a lack of association between HMGA1 expression and the morphologic grading of ductal breast carcinomas.
HMGA1 expression correlates with Her2/neu expression and ER and PR status

Subsequently, we have evaluated the relationship between HMGA1 expression and several indicators of breast carcinoma invasion (Her2/neu amplification, ER and PR status). As reported in Table 1, HMGA1 expression was significantly associated with that of Her2/neu (p=0.004). Indeed, HMGA1 and Her2/neu staining are positively correlated: low levels of HMGA1 (26.5±29.7) are associated with a weak Her2 staining (0+1), whereas high levels of HMGA1 (32.2±30.9) are correlated with an intense Her2 staining (2+3).

As far as the correlation of HMGA1 expression with the endocrine status of the breast carcinoma samples is concerned, a significant association has been found with the retention of PR (p=0.003) and loss of ER expression (p=0.007). Indeed, expression of PR was found positively correlated with high levels of HMGA1 (29.3±31.1) in breast carcinomas (Table 1), while ER expression was found negatively correlated with high levels of HMGA1 (31.1±30.9, Table 1). Therefore, in human breast carcinomas the over-expression of HMGA1 is positively and negatively correlated with PR and ER status, respectively (Table 1). Representative cases of breast carcinomas expressing HMGA1, PR and ER are reported in Figure 2 (Figure 2A, B, C, D).
DISCUSSION

The aim of our study was to verify whether HMGA1 protein might be an indicator for the diagnosis and prognosis of human breast carcinoma. Therefore, we investigated by immunohistochemistry the expression of the HMGA1 protein in \( n=1202 \) breast carcinoma tissues combined on a TMA and compare it with multiple clinicopathological parameters.

Consistently with our previous studies showing that HMGA1 protein over-expression was found in 60% of ductal carcinomas and in almost all of the lobular carcinomas (Chiappetta et al. 2004), we found that the HMGA1 protein resulted over-expressed in the breast carcinoma specimens compared to the breast normal control tissues and, in addition, a differential HMGA1 protein expression was found between all the carcinoma samples analysed. No association was found between HMGA1 expression and histological grade of ductal carcinomas. Conversely, we have observed an inverse trend between the over-expression of HMGA1 and the \( pN \) stage (if we exclude the \( pN3 \) value).

A strong association was, instead, observed between HMGA1 over-expression and Her2/neu amplification in breast carcinoma. This result is in line with our previously published data, obtained on a shorter series of human breast carcinomas where HMGA1 over-expression correlated with the amplification of ErbB2 (Chiappetta et al. 2004). Her2 is a transmembrane protein with substantial homology to epidermal growth factor receptor. Amplification of this gene coupled with resultant over-expression of the protein occurs in about 25% of human breast cancers. It has been reported that amplification of c-ErbB2 is associated with a fast proliferation and a poor prognosis of breast cancer (Vogelstein and Kinzler, 1994). Therefore, HMGA1 over-expression could not only represent a predictor of prognosis and outcome, as exerted by amplification of Her2, but it can be also predicted a functional link with the amplification of the Her2 gene. In fact, several studies have demonstrated an accumulation of amplifications of different genomic regions in certain breast cancers considered to exhibit an “amplifier” phenotype (Al-Kuraya et al. 2004, Courjal et al. 1997). Therefore, we could hypothesize that that HMGA1 over-expression may contribute to the accumulation of Her2/neu amplifications, or alternatively that the activation of the ErbB2 transduction pathway may lead to increased HMGA1 protein synthesis.
Likewise, we found a strong relationship between the HMGA1 over-expression and the presence or absence of PR and ER, respectively. It is well known that breast carcinomas not expressing ER are insensitive to estrogen antagonists such as tamoxifen and, as consequence, these carcinomas are insensitive to hormonal treatments and display a negative outcome (Teschendorff et al. 2007). These findings are consistent with HMGA1 over-expression seems associated with a highly malignant phenotype, also representing a poor prognostic index since HMGA1 over-expression often correlates with the presence of metastasis, and with a reduced survival, as it has been extensively reported elsewhere (Abe et al. 2003, Meyer et al. 2007).

On the other hand, the positive correlation between the HMGA1 over-expression and the presence of the PR (Table 1), apparently would link the expression of HMGA1 to a better prognosis, since PR+ breast carcinoma patients respond to the hormonal treatment, even though there are reports indicating that breast cancer patients with ER-/PR+ tumours arose primarily premenopausal and in younger people (Rakha et al. 2007, Rhodes et al. 2009, Cserni et al. 2011). Although these ER-/PR+ patients are generally considered candidates for endocrine therapy, they gain less benefit from adjuvant tamoxifen treatment than ER+/PR+ patients and ER-/PR+ patients <55 years old were found to have significantly worse survival than younger ER+/PR+ patients (Yu et al. 2008). Therefore, considering the outcome, the PR expression needs to be evaluated together with the ER status, whose expression is negatively related to that of HMGA1. Since HMGA1 is a chromatin protein able to up- or down-regulate the expression of important cancer-related genes (Fusco and Fedele, 2007), it could be envisaged that HMGA1 proteins could be able to modulate the expression of both ER and PR expression acting on its promoter. This could have important prognostic significance, since the over-expression of HMGA1 could influence the hormonal responsiveness and, therefore, the outcome of breast cancer patients.

In conclusion, taken together, our findings indicate that the level of HMGA1 expression is related to an unfavourable breast cancer phenotype and poor prognosis, as supported by its strong association with the Her2/neu, PR and ER status that could explain, at least in part, the different behaviour of the human breast carcinoma over-expressing HMGA1. These findings further support the HMGA1 as
an appropriate target for the therapy of human cancer, as suggested by numerous *in vitro* and *in vivo* studies modulating its expression in cancer cells.
ACKNOWLEDGEMENTS

This work was supported by grants from the Associazione Italiana per la Ricerca sul Cancro (AIRC), Progetto Strategico Oncologia, Consiglio Nazionale delle Ricerche, the Ministero dell'Università e della Ricerca Scientifica e Tecnologica (MIUR), PNR-CNR Aging Program 2012-2014. We thank the Associazione Partenopea per le Ricerche Oncologiche (APRO) for its support.

CONFLICT OF INTEREST STATEMENT

The authors have declared that no conflict of interest exists.
REFERENCES


Clinical Cancer Research. Vol. 10, 7637–7644, November 15, 2004


Miecznikowski et al. BMC Cancer 2010, 10:573


Chia-Chen Liu, Julie Prior, David Piwnica-Worms, and Guojun Bua. LRP6 overexpression defines a class of breast cancer subtype and is a target for therapy. approved January 28, 2010 (received for review October 1, 2009) 5136–5141 | PNAS | March 16, 2010 | vol. 107 | no. 11


Figures and Tables:

Table 1: Patient characteristics and association of HMGA1 with clinico-pathological features of breast cancer patients. Raw, continuous scores (% positive staining) are used.

<table>
<thead>
<tr>
<th>Clinico-pathological feature</th>
<th>HMGA1 Mean (%) ± SD, median (%)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histological subtype</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ductal carcinoma (n=963)</td>
<td>27.4±30.4 12.5</td>
<td>0.883</td>
</tr>
<tr>
<td>Other (n=375)</td>
<td>26.9±28.4 15</td>
<td></td>
</tr>
<tr>
<td>pT stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pT1 (n=447)</td>
<td>28.4±30.2 15</td>
<td>0.402</td>
</tr>
<tr>
<td>pT2 (n=656)</td>
<td>27.2±29.9 11.3</td>
<td></td>
</tr>
<tr>
<td>pT3 (n=84)</td>
<td>26.9±27.9 18.8</td>
<td></td>
</tr>
<tr>
<td>pT4 (n=134)</td>
<td>24.7±29.7 10</td>
<td></td>
</tr>
<tr>
<td>pN stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pN0 (n=598)</td>
<td>29.0±30.4 20</td>
<td>0.293</td>
</tr>
<tr>
<td>pN1 (n=452)</td>
<td>26.4±29.5 10</td>
<td></td>
</tr>
<tr>
<td>pN2 (n=105)</td>
<td>24.7±28.3 10</td>
<td></td>
</tr>
<tr>
<td>pN3 (n=27)</td>
<td>34.4±34.5 20</td>
<td></td>
</tr>
<tr>
<td>Her2/neu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staining 0+1 (n=1059)</td>
<td>26.5±29.7 10</td>
<td>0.004</td>
</tr>
<tr>
<td>Staining 2+3 (n=201)</td>
<td>32.2±30.9 20</td>
<td></td>
</tr>
<tr>
<td>ER status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0% (n=322)</td>
<td>31.1±30.9 20</td>
<td>0.007</td>
</tr>
<tr>
<td>&gt;0% (n=978)</td>
<td>26.2±29.5 10</td>
<td></td>
</tr>
<tr>
<td>PR status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0% (n=603)</td>
<td>23.6±28.0 10</td>
<td>0.003</td>
</tr>
<tr>
<td>&gt;0% (n=402)</td>
<td>29.3±31.1 20</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Immunohistochemical analysis of a human breast cancer TMA for HMGA1 protein expression. Examples of HMGA1 immunohistochemistry results in breast carcinomas: negative (A), weak (B), moderate (C), and strong (D) staining.
Figure 2.
Representative IHC staining of human ductal breast cancers.

Figure 2. Immunohistochemical analysis of HMGA1 expression in breast ductal carcinomas. ER- breast carcinoma (A). ER+ breast carcinoma (B). PR- breast carcinoma (C). PR+ breast carcinoma (D).
6.4 Effect of EpCAM, CD44, CD133 and CD166 expression on patient survival in tumours of the ampulla of Vater.

My contribution to this work:
- Immunohistochemical analysis of whole tissue and microarray sections;
- Statistical comparison and survival analysis;
- Data analysis and manuscript writing;
Effect of EpCAM, CD44, CD133 and CD166 expression on patient survival in tumours of the ampulla of Vater

Salvatore Piscuoglio,1 Frank S Lehmann,2 Inti Zlobec,1 Luigi Tornillo,1 Wolfgang Dietmaier,3 Arndt Hartmann,4 Peter H Wünsch,5 Fausto Sessa,6 Petra Rümmele,3 Daniel Baumhoer,1 Luigi M Terracciano1

ABSTRACT
Background Carcinomas of the Vaterian system are rare and presumably arise from pre-existing adenomas. According to the cancer stem cell (CSC) hypothesis, only a small subset of tumor cells has the ability to initiate and develop tumor growth. In colorectal cancer, CD44, CD133, CD166 and EpCAM have been proposed to represent CSC marker proteins and their expression has been shown to correlate with patient survival.

Aims To evaluate a potential role of these CSC proteins in tumors of the ampulla of Vater, we investigated their expression in 175 carcinoma, 111 adenoma and 152 normal mucosa specimens arranged in a Tissue Microarray format.

Materials and methods Membranous immunoreactivity for each protein marker was scored semi-quantitatively by evaluating the number of positive tumor cells over the total number of tumor cells. Median protein expression levels were used as cut-off scores to define protein marker positivity. Clinical data including survival time were obtained by retrospective analysis of medical records, tumor registries or direct contact.

Results The expression of all evaluated marker proteins differed significantly between normal mucosa, adenoma and carcinoma samples. In all markers, we found a tendency towards more constant expression from normal to neoplastic tissue. EpCAM expression was significantly correlated with better patient survival. The increased expression of CD44s, CD166 and CD133 from normal mucosa samples to adenoma and carcinoma was linked to tumor progression. However, there was no statistically significant correlation with survival.

Conclusion Our findings indicate, that in ampullary carcinomas, loss of expression of EpCAM may be linked to a more aggressive tumor phenotype.

INTRODUCTION
The ampulla of Vater combines the terminal and common segment of the bile and pancreatic duct before they enter the duodenum.1 Carcinomas originating from this complex anatomical unit are uncommon and have an incidence of approximately four to six cases per million population.2,3 Carcinomas of the papilla of Vater, defined as junction of the biliary, and pancreatic ducts within the duodenum account for 6%–20% of all peripancreatic tumours4 and represent 10%–50% of all cancers resected by pancreaticoduodenectomy.5 They can be sited in the ampulloduodenal part of the papilla of Vater, which is lined by intestinal mucosa. They also can develop in deeper parts of the ampulla, which are lined by pancreatobiliary duct mucosa. Clinically, tumours of the ampulla of Vater are rapidly detected due to biliary outflow obstruction.6,7 Early symptoms as well as differences in tumour biology are held responsible for their favourable clinical outcome (median survival 50–50 months, 5-year survival rate 21%–64%).8,9 Histologically, intestinal, pancreatobiliary, intestinal-mucinous, invasive papillary and poorly differentiated subtypes can be distinguished.10 The subtypes differ in several clinical and histological aspects including cell type-specific markers, oncogene expression, modes of tumour spread as well as extent and interaction with the extracellular matrix.11 Most authors agree that local spread of the tumour (T stage) is the only significant and independent prognostic factor for this cancer, whereas the predictive value of tumour grade and lymph node metastases is still debated.12,13 More recent research data suggest that the prognosis of ampullary cancer may be related to the histological differentiation in intestinal or pancreatobiliary types.14 In the last years, several molecular markers have been proposed as additional prognostic factors. However, most of these studies have yielded conflicting results and have not been still validated by other reports.15–19 Several sources of discrepancy between different reports have been acknowledged mainly due to non-standardised assays often performed on underpowered patient samples that are too small to enable meaningful conclusions to be drawn. Therefore, there is undoubtedly a need for additional prognostic markers for such neoplasia. Recent findings support the concept that cells with the properties of stem cells are integral to the development and perpetuation of several forms of human cancer.20–21 Cancer stem cells (CSCs) have low replicative ability, multipotency and resistance to apoptosis and are responsible for tumour development.22 In the different types of digestive tumours, different sets of markers have emerged as the most useful for the identification of CSC. In particular, in intestinal as well as in pancreatic cancer, some markers including CD44, EpCAM, CD166 and CD133 have been indicated as possible CSCs markers. Furthermore, in colorectal cancer, we have shown that their expression inversely correlated with patient survival.23 However, conflicting results have been
reported about the role of some putative CSC markers in gastrointestinal tract tumours. In particular, contradictory findings have been reported about the association of CD44, in particular of its v6 splicing variant, and tumour progression.\textsuperscript{24–26} Furthermore, while CD133 molecule was initially identified as a reliable CSC marker in human colorectal cancers,\textsuperscript{27,28} a subsequent study has shown that in both mouse and human colorectal cancers, CD133 expression is not restricted to rare cell subsets, but it is detectable in a large majority of tumour cells, irrespective of their tumorigenicity.\textsuperscript{29} Because of the lacking studies dealing with CSC markers in ampullary tumours, the aim of this study was to elucidate the expression and the prognostic role of CD133, CD166, CD44s, EpCAM expression in ampullary tumours by using a tissue microarray (TMA) including 175 carcinoma, 111 adenoma and 152 normal mucosa specimens of the papilla of Vater.

**MATERIAL AND METHODS**

**Ethics**

The study has been approved by the institutional review board of the Department of Pathology, University of Basel, Switzerland. All the analyses were performed according to the ethical standards required by each local ethic committee.

**Patients’ characteristics and tissue samples**

Patients’ characteristics have been previously described by our study group.\textsuperscript{30–31} Briefly, the files of the Institute of Pathology, University Hospital Basel (Switzerland), the Institute of Pathology, University of Regensburg (Germany), the Institute of Pathology Nuernberg and the Anatomic Pathology Unit, Department of Human Morphology, University of Insubria, Varese (Italy), were searched for adenomas or carcinomas of the ampulla of Vater over the period from 1982 to 2005. In total, 175 carcinoma, 111 adenoma and 152 normal mucosa samples were retrieved. Sufficient paraffin-embedded tissue for TMA construction was available in all cases. The male-to-female ratio was 3:2; mean age at diagnosis was 63 years (range 15–81 years). To our knowledge, no case was associated with Familial adenomatous polyposis (FAP).

**TMA construction**

TMAs were constructed from formalin-fixed and paraffin-embedded specimens using a custom-built instrument (Beecher Instruments, Silver Spring, Maryland, USA) as previously described.\textsuperscript{30–31} Briefly, H&E-stained sections were obtained from each selected primary block (donor block) to define representative tissue regions. Core biopsies (0.6 mm cylinders) were taken from the selected tissue regions and then transferred to a paraffin recipient block. The resulting TMA was cut into 4 µm sections, which were used for immunohistochemistry. The number of punches per patient ranged from one to three for both normal tissue and carcinoma and from one to five for adenomas. If more than one punches was obtained, the additional punches were taken from different representative blocks.

**Histology and immunophenotyping**

All tumours were classified according to the guidelines of the Armed Forces Institute of Pathology using only H&E stains.\textsuperscript{10} Mild dysplasia was designated as low-grade dysplasia, whereas moderate and severe dysplasia was considered as high-grade dysplasia. Carcinomas histologically indistinguishable from colorectal carcinomas were classified as intestinal types, whereas carcinomas showing a dense desmoplastic stroma surrounding small glands or solid nests of tumour cells were referred to as the pancreaticobiliary subtype. Invasive papillary carcinomas typically formed papillary and micropapillary structures in their invasive component and poorly differentiated adenocarcinomas lacked histologic features of glandular or other differentiation. Additionally, an intestinal-mucinous subtype, characterised by any mucinous differentiation, was defined.

For immunohistochemistry, sections were pre-treated with CC1 (Ventana Medical Systems, Tucson, Arizona, USA) and incubated with primary antibodies against CD133, CD44s, CD166 and EpCAM (table 1). Staining procedures were performed on a Benchmark immunohistochemistry staining system (Ventana Medical Systems) using iVIEW-DAB as chromogen.

Membranous immunoreactivity for each protein marker was scored semiquantitatively by evaluating the number of positive tumour cells over the total number of tumour cells. Scores were assigned using 5% intervals and ranged from 0% to 100%. All tissues were scored by an experienced pathologist (LT), blinded to clinicopathological information. To define interobserver agreement, all samples were examined independently by a second pathologist (DB).

**Statistical analysis**

Statistical correlations between categorical variables were tested using a χ² or Fisher exact test, where appropriate. Differences in patient survival were demonstrated using the Kaplan–Meier method and analysed using the log-rank test in univariate analysis. All tests were two sided. p Values <0.05 were considered statistically significant. Cut-off scores were selected by evaluating the receiver operating characteristic curves for each protein marker and the end-point survival. The point on the curve with the shortest distance to the coordinate (0, 1) was selected as the threshold value to classify cases as ‘positive/overexpressing’ or ‘negative/loss’.\textsuperscript{32} Analysis was performed using SAS V9.1 (SAS Institute).

**RESULTS**

**Patients’ characteristics and tissue samples**

Overall, 175 carcinoma, 111 adenoma and 152 normal mucosa samples were retrieved. In patients with carcinomas, we found 19 pT1 (11%), 59 pT2 (34%), 65 pT3 (56%) and 15 pT4 (7%) tumours as well as 17 G1 (10%), 82 G2 (47%) and 55 G3 (31%) cases (no data concerning T stage and grading in 21 samples). Seventy-three (42%) carcinoma patients were node positive and two (1%) had haemotogenous metastases at initial diagnosis. Full clinical data including survival time were available in 133 patients with ampullary carcinoma (76%). Patients were studied up to 164 months after operation. Median follow-up time was 56 months.

The histological classification of 175 ampullary carcinomas identified 85 intestinal types, 42 pancreaticobiliary types, 23 poorly differentiated adenocarcinomas, 16 intestinal-mucinous types and nine invasive papillary types.

**Table 1** Primary antibodies against CD133, CD44, CD166 and EpCAM

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Dilution/detection</th>
<th>Pretreatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD44 (Dako, DF1485)</td>
<td>1:50/BOND</td>
<td>BOND ER2</td>
</tr>
<tr>
<td>CD133 (cell signalling, C24B9)</td>
<td>1:100/BOND</td>
<td>Steam 120°C, pH8</td>
</tr>
<tr>
<td>CD166 (Novocastra, M05/07)</td>
<td>1:200/BOND</td>
<td>BOND ER2</td>
</tr>
<tr>
<td>EpCAM (Novocastra, VU-1D9)</td>
<td>1:200/BOND</td>
<td>BOND ER2</td>
</tr>
</tbody>
</table>
Histological grading was evaluable in 76 of 111 (68%) adenomas and disclosed low-grade dysplasia in 57 of 76 (75%) and high-grade dysplasia in 19 of 76 (25%) cases. All adenomas demonstrated tubular or tubulovillous architecture. Seventy-eight of 111 (70%) adenoma samples were derived from patients with coexisting carcinoma.

Immunophenotyping

Tissue samples of ampullary carcinoma patients expressing CD44, CD133, CD166 and EpCAM are shown in figure 1. Moreover, in table 2, the distribution of the different biomarkers across different diagnostic categories is shown.

The expression of all marker proteins differed significantly between carcinoma, adenoma and normal mucosa samples (table 3). We have also evaluated the positivity in the two principal histologic types (intestinal type vs pancreatobiliary type). EpCAM was significantly more expressed in intestinal type (table 4).

We have tried also to evaluate if there is some difference between adenomas without coexisting carcinoma and adenomas without coexisting carcinomas. Only CD44 was significantly more expressed in adenomas with coexisting carcinomas (p=0.045).

No difference was found between low-grade and high-grade adenomas (data not shown).

Survival

Five-year survival (95% CI) was 45.2 (34 to 56) in EpCAM-positive versus 28.2 (11 to 48) in EpCAM-negative patients (p<0.05). EpCAM was not an independent prognostic factor after adjusting for pT and pN stages. Survival curves of both patient groups using the Kaplan–Meier method are demonstrated in figure 2.

DISCUSSION

Tumours of the papilla of Vater are a relatively rare neoplastic entity that came into focus in recent years. Significant overlap exists in phenotypic and molecular characteristics between ampullary and colorectal carcinomas. As in colorectal cancer, the development of ampullary carcinoma from adenomas as precancerous lesions has been well documented, and studies investigating molecular alterations associated with the proposed adenoma–carcinoma sequence have been also performed, including our group. However, still missing is a comprehensive analysis of the expression of putative CSC markers in very large groups of patients, amenable to detailed statistical analysis. Moreover, the prognostic significance of the co-expression of multiple CSC markers within the same tumour has not been evaluated so far.

This is the first systematic study assessing the prognostic value of four CSC markers, namely EpCAM, CD44, CD133 and CD166, in a large series of patients with ampullary tumours.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Distribution of biomarkers between different diagnostic categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of cases within each expression category</td>
</tr>
<tr>
<td></td>
<td>0%</td>
</tr>
<tr>
<td>Normal</td>
<td></td>
</tr>
<tr>
<td>CD133</td>
<td>9</td>
</tr>
<tr>
<td>CD44</td>
<td>67</td>
</tr>
<tr>
<td>CD166</td>
<td>81</td>
</tr>
<tr>
<td>EpCAM</td>
<td>7</td>
</tr>
<tr>
<td>Adenoma</td>
<td></td>
</tr>
<tr>
<td>CD133</td>
<td>10</td>
</tr>
<tr>
<td>CD44</td>
<td>25</td>
</tr>
<tr>
<td>CD166</td>
<td>58</td>
</tr>
<tr>
<td>EpCAM</td>
<td>0</td>
</tr>
<tr>
<td>Cancer</td>
<td></td>
</tr>
<tr>
<td>CD133</td>
<td>16</td>
</tr>
<tr>
<td>CD44</td>
<td>38</td>
</tr>
<tr>
<td>CD166</td>
<td>84</td>
</tr>
<tr>
<td>EpCAM</td>
<td>2</td>
</tr>
</tbody>
</table>
The overexpression of EpCAM was significantly correlated with better survival time. The increasing expression of CD44, CD166 and CD133 from normal mucosa to adenoma and carcinoma was linked to tumour progression. However, there was no statistically significant correlation with survival.

EpCAM is a glycosylated, 30–40 kDa type I membrane protein, which is expressed in a variety of human epithelial tissue cancers, as well as in progenitor and stem cells. It is composed of an extracellular domain with epidermal growth factor and thyroglobulin repeat-like domains, a single transmembrane domain and a short 26 amino acid intracellular domain called EpICD. In normal cells, EpCAM is predominantly located in intercellular spaces, where epithelial cells form very tight junctions. Therefore, on normal epithelia, it is sequestered and may be much less accessible to antibodies than in cancer tissue, where it is homogeneously distributed on the cell surface. Furthermore, EpCAM is part of the signature of cancer-propagating cells in numerous solid tumours as well as in normal progenitor and stem cells.  

EpCAM was one of the first tumour-associated antigens identified in the late 1970s. Systematic analysis of EpCAM expression for intensity and frequency showed that EpCAM is expressed on essentially all human adenocarcinoma, on certain squamous cell carcinoma, on retinoblastoma and on hepatocellular carcinoma.  

Importantly, EpCAM is part of the signature of cancer-propagating cells in numerous solid tumours and of normal progenitor and stem cells. The controversial biological role of EpCAM has recently been discussed by van der Gun et al. It is of interest that EpCAM overexpression has been associated with both decreased and increased survival time. EpCAM exerts different effects on cell adhesion, either promoting or preventing metastasis. The correlation of EpCAM expression and poor survival has been described in several tumour types, including invasive breast cancer, urothelial carcinoma of the bladder, gallbladder carcinoma and squamous cell carcinoma of the oesophagus. In different tumours, studies on EpCAM-directed immunotherapeutic therapies are currently in clinical development. Therapeutic trials of monoclonal antibodies directed against EpCAM have shown that they may induce antibody-based cellular cytotoxicity by adhering to cytokines such as interleukin 2 or complement-based cytotoxicity by activating T cytotoxic cells.  

Several reasons for the discrepancies between our results and previous studies can be hypothesised including differences in sample size (power for detecting prognostic differences), methodology (TMA vs whole tissue sections), different clones of antibody and, most importantly, the choice of cut-off scores for the definition of positive staining or staining intensity.

EpCAM is intensely used as a therapeutic target for antibody-based approaches. Future development of EpCAM-directed therapeutics may profit from newly identified functions of EpCAM as mitogenic signal transducer in various ways. An important insight is that EpCAM is apparently needed to maintain distinct cancer cell attributes and, potentially, the CSC phenotype as well. This function can reduce the risk of immune escape by loss of EpCAM target expression from cancer cells. EpCAM-directed therapies may be selective for those cancer cells with the strongest negative impact on prognosis and for cancer-propagating subsets of malignant cells. Discrepant results have also been reported regarding the effect of CD44 gene or protein and its splice variants on survival in tumour patients and it is an important receptor that binds hyaluronan (HA). CD44 has previously been considered to be a marker of tumour invasiveness and metastasis. Only recently, it has been described as putative colorectal CSC marker.
However, CD44 does not seem to belong to the group of genes, such as OCT4 and NANOG, that are central for maintaining stem cell characteristics. Nonetheless, two connections between CD44 and genes that regulate stem cell characteristics have been described. First, CD44 is a target of the WNT pathway. Loss of adenomatous polyposis coli (APC) function leads to the constitutive activation of β catenin, a constituent of the WNT signalling pathway. CD44s and CD44v6 expression is restricted to the intestinal crypts in non-transformed tissue, but both CD44 isoforms are strongly overexpressed in dysplastic crypts and adenomas in humans and mice with mutant APC.47 Second, HA–CD44 binding promotes protein kinase C (PKC) activation and this increases NANOG phosphorylation and translocation to the nucleus. Here, it associates with Drosha and an RNA helicase, p68, leading to the transcription of the oncogenic microRNA (miRNA) miR-21 and a reduction in the expression of the tumour suppressor programmed cell death 4. These events initiate the upregulation of the inhibitor of apoptosis (IAP) proteins and multidrug resistant protein I (MDR1). CD44, in turn, associates with and stabilises MDR1 expression.48 This could be one mechanism through which CD44 contributes to stem cell resistance to chemotherapy, as MDR1 exports several drugs from cells.

In several gastrointestinal tumours, including colorectal cancer,23 stromal25 as well as neuroendocrine tumours,50 loss of CD44 expression has been associated with disease progression and reduced survival. In our study, CD44 expression was more frequently associated with adenomas with coexisting carcinomas.

Our study is also the first to evaluate the prognostic impact of CD166 in ampullary carcinoma. We found an increasing expression of CD166 from normal tissue to carcinoma, suggesting that the increased expression of this marker might be linked to tumour progression. Our data are supported by a recent study in colorectal cancer patients, where a similar increasing expression of CD166 from normal to neoplastic tissue has been described by Weichert et al.51

We speculate that our findings of decreased rather than increased expression of membranous EpCAM expression and its association with features of tumour progression are mainly a consequence of its cell adhesion function. In colorectal cancer, Kojima et al.42 and Horst et al.28 reported a significant correlation of increased CD133 expression and poor clinical outcome. In contrast, in a study on non-small-cell lung cancer, CD133 expression was not a prognostic factor for survival.44 Consistent with the latter study, we found no significant impact of CD133 on survival in our series of ampullary cancer patients.

In summary, we have provided evidence that in ampullary carcinoma, loss of expression of EpCAM, but not of CD44, CD133 or CD166, is linked to poor survival.

Competing interests None.

Patient consent If yes, who signed it?

Ethics approval Ethische Kommission Beider Basel.

Contributors SP, FSL and DB chose the antibodies, collected the data, wrote the introduction and the discussion. LMT and DB performed the immunohistochemical staining. IZ performed the statistical evaluation. LT, LMT, PR, FS, PHW and WD revised and reclassified the tumours. LMT and LT revised the manuscript.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

good, the bad or the ugly.


6.5 MAGE-A10 is a nuclear protein frequently expressed in high percentages of tumor cells in lung, skin and urothelial malignancies.

My contribution to this work:
- Immunohistochemical analysis of whole tissue and microarray sections;
- Data analysis and statistical comparison;
MAGE-A10 is a nuclear protein frequently expressed in high percentages of tumor cells in lung, skin and urothelial malignancies

Elke Schultz-Thater1*, Salvatore Piscuoglio2*, Giandomenica Iezzi3, Clémentine Le Magnen1, Paul Zajac1, Vincenza Carafa3, Luigi Terracciano2, Luigi Tomillo3* and Giulio C. Spagnoli1*  

1Institute for Surgical Research and Hospital Management and Department of Biomedicine, University of Basel, Basel, Switzerland  
2Institute of Pathology, University of Basel, Basel, Switzerland

MAGE-A10 is a highly immunogenic member of the MAGE-A family of cancer/testis-associated antigens (C/T TAAs). Studies performed with broadly reactive antibodies have helped to initially characterize this TAA. However, no specific reagents have been developed so far, thus preventing a thorough analysis of its expression in healthy and tumoral tissues. We have produced MAGE-A10 gene product in soluble recombinant form, and we have used it to generate specific monoclonal antibodies (mAbs). One of these reagents, recognizing an epitope located at the COOH terminus of the MAGE-A10 gene product, was used to stain a multitumor tissue microarray comprising more than 2,500 paraffin-embedded specimens including healthy tissues, benign tumors and malignancies of different histological origin. MAGE-A10 protein was identified as an intranuclear protein of an apparent molecular weight of 70 kDa, expressed in normal spermatogonia and spermatocytes but in no other healthy tissue. Most importantly, this C/T TAA appears to be expressed in high (>50%) percentages of cancer cells from a number of malignancies, including lung, skin and urothelial tumors. Unexpectedly, high expression of MAGE-A10 TAA at the protein level was also detectable in gynecological malignancies and stomach and gall bladder cancers. The characterization of MAGE-A10-specific reagents might set the stage for the development of targeted active immunotherapies by clarifying potential indications and by allowing the selection of patients eligible for treatment and the monitoring of its effectiveness.

MAGE-A tumor-associated antigens of the cancer/testis family are expressed in a very limited number of healthy tissues typically including spermatogonia, and, for some of them, thymus and placenta. In contrast, they are expressed in a large variety of malignancies derived from diverse tissues, including, among others, melanoma, lung cancers, prostate cancers, breast cancers, hepatocellular carcinoma and head and neck cancers.1,2

In these tumors, their expression can be used for diagnostic purposes or for the identification of patients potentially benefited of targeted immunotherapies. Quantitative real-time PCR (qRT-PCR) is currently used for the detection, at the gene level, of the expression of MAGE-A tumor-associated antigens. On the other hand, only a limited number of reagents are available to detect the corresponding proteins.3

MAGE-A10 is probably the most immunogenic antigen of the MAGE-A family,4–8 and, therefore, it represents a potentially highly attractive target of active specific immunotherapies. This antigen, expressed in the form of a 72-kDa nuclear protein, has been identified thanks to broadly reactive monoclonal antibodies (mAbs) recognizing it together with other members of the MAGE-A family.3 However, no exquisitely specific reagents have been developed so far, thus preventing a precise identification of cells expressing it in healthy and cancerous tissues and, in particular, in paraffin-embedded specimens. This information is of critical relevance in the selection of patients potentially eligible for targeted active immunotherapies and for the monitoring of their effectiveness.

Upon expression cloning, we have produced MAGE-A10 in soluble form and we have used it to generate highly specific mAbs. Here, we report that MAGE-A10-specific mAbs are able to identify their target antigen in the nuclei of tumor cells in paraffin-embedded clinical samples. Staining of a multitumor array including more than 2,000 specimens indicates that MAGE-A10 cancer/testis tumor-associated antigen (C/T TAA) is highly expressed in lung, skin and urothelial malignancies.

Key words: cancer/testis tumor-associated antigens, MAGE-A10, immunohistochemistry, tissue microarray, cancer immunotherapy

Grant sponsor: SNF; Grant numbers: 320030-120320, 3100A0-122235/1  
*E.S.-T., S.P., L.T. and G.C.S. contributed equally to this work  
DOI: 10.1002/ijc.25777  
History: Received 30 Jun 2010; Accepted 14 Oct 2010; Online 15 Nov 2010  
Correspondence to: Giulio C. Spagnoli, Institute for Surgical Research and Hospital Management and Department of Biomedicine, University of Basel, 20 Hebelstrasse, 4031 Basel, Switzerland, Tel.: +41-61-265-2378, Fax: +41-61-265-3990, E-mail: gspagnoli@uhbs.ch
Material and Methods

Cell lines
MZ-2 and A375 cell lines are gifts of Dr. Rimoldi (Ludwig Institute, Lausanne, Switzerland). RE cell line is a gift of Dr. Siegrist, formerly at the University of Basel, whereas WM115 cell line was obtained from American Type Culture Collection (Rockville, MD). SK-Mel-37 cell line is a gift of Dr. Jungbluth (Ludwig Institute at Memorial Sloan Kettering Cancer Center, New York, NY). Na8 cell line is a gift of Dr. Jotereau (Nantes, France). All cell lines were cultured in RPMI 1640 medium, supplemented with 10% FCS, nonessential amino acids, glutamine and antibiotics (all from Invitrogen, Basel, Switzerland).

Preparation of MAGE-A10 fusion protein
MAGE-A10 entire gene was PCR amplified from cDNA derived from the SK-Mel-37 melanoma cell line by using the following primers, allowing the cloning into a suitable expression vector and including EcoRI and Hind III restriction sites:
forward: 5’ CCGGAATTCCTGGACTCTCAACGGTCAGCCG
reverse: 5’ CCAAAGCTTATTCAGGGTAGGAGAA 3’

A 1110-bp band was excised and inserted into pET-32a vector (Novagen, Madison, WI) allowing inducible expression of inserted genes in the form of fusion proteins containing thioredoxin and a histidine tail. The plasmid was used to transform BL21(pLysS) E. coli strain. After a 4-hr induction in the presence of IPTG (1 mM final concentration), bacterial cultures were lysed and recombinant proteins were purified under native conditions upon binding to nickel resins (Ni-NTA, Qiagen, Basel, Switzerland) and eluted in the presence of 250 mM imidazole. Production and purification of the recombinant proteins were monitored by SDS-PAGE and Coomassie blue staining (Fig. 1a). Additional recombinant C/T TAAs was similarly produced in E. coli.

Production of monoclonal antibodies
BALB/c mice were repeatedly injected i.p. at 2-week intervals with 100 µg Ni-purified material containing MAGE-A10 protein, in the presence of Sigma adjuvant system (Sigma, Buchs, Switzerland). Three days after a last injection, animals were sacrificed, and fusions were carried out according to standard methods. Screening of HAT-resistant hybridomas was performed by ELISA.

Detection of MAGE-A10 gene expression
Total cellular RNA was extracted from the cell lines under investigation (see below), reverse transcribed and tested in qRT-PCR assays in the presence of primers and probes specific for β-actin-positive control gene and of the following MAGE-A10-specific reagents:
forward: 5’ TCAGGAGGAGACAAGGTCAAGA 3’
reverse: 5’ GGAGGAGTTGGAAGGACGACTT 3’

Specific gene expression was quantitated by using the 2-ΔΔCT method on data normalized by using β-actin as reference gene. Results were expressed as ratio to SK-Mel-37 reference cell line.

Epitope mapping
The FlTrx random peptide library (Invitrogen, Basel, Switzerland) composed of 1.77 × 10^8 primary clones of E. coli with a dodecamer peptide sequence inserted within the Thioredoxin (TrxA) active site loop was used to identify epitopes recognized by MAGE-A10-specific mAb, as previously detailed. Briefly, 2 ml of the FlTrx library was induced in the presence of IMC medium containing ampicillin and 100 µg/ml tryptophan at 25°C. Bacteria were then “panned” on 60-mm tissue culture plates (Becton Dickinson, Franklin Lakes, NJ) pretreated with the mAb under investigation (50 µg/ml in sterile water). After the removal of unbound cells, bacteria were amplified o/n in IMC medium at 25°C. After five rounds of panning and culture, E. coli were streaked onto RMG medium plates containing 100 µg/ml ampicillin and incubated overnight at 30°C. Individual clones tested positive in Western blot were selected and amplified o/n at 30°C in RM medium containing ampicillin. DNA was isolated using the Nucleospin kit (Macherey-Nagel, Oensingen, Switzerland), and plasmids were sequenced by using the FlTrx forward or Rsr reverse sequencing primers.

Tissue microarrays and immunohistochemistry
Tissue microarray (TMA) construction and staining were described in detail elsewhere. Briefly, formalin-fixed and paraffin-embedded tissues were obtained from the archives of the Institute of Pathology at the University of Basel. The multitissue TMA used in our study comprised a total of 2,587 samples including 218 normal tissues, 518 benign tumors and 1,851 malignancies from more than 100 different tumoral tissues. Slides were independently scored by three members of the team, including two experienced pathologists (LTc and LTt). Samples were considered positive if at least 5% of cells showed evidence of staining of moderate or strong intensity.

Results
Production of recombinant MAGE-A10
MAGE-A10 complete open reading frame was PCR amplified from SK-Mel-37 cell line cDNA and cloned into the pET 32a-inducible expression vector. Upon IPTG treatment, the TAA was produced in soluble form within a fusion protein, inclusive of thioredoxin and a polyhistidine tail. After purification on nickel columns, this protein was detectable in Coomassie blue-stained gels with an apparent molecular weight (MW) of 79 kDa (Fig. 1a). This material was used to immunize mice and screen hybridomas. Similarly produced soluble thioredoxin (TrxA) served as negative control in screening procedures.

Generation of MAGE-A10-specific mAbs recognizing recombinant and native gene products
A number of mAbs (>20) appeared to recognize recombinant MAGE-A10 protein in ELISA assays (data not shown). To verify their capacity to identify the native protein, they were
tested in "Western" blot assays on lysates from SK-Mel-37 cell line, originally utilized for the cloning of the MAGE-A10 gene (see above). Although all reagents recognized the positive control 79-kDa recombinant fusion protein, notably, 15 of 20 mAbs identified a single band of an apparent MW of 70 kDa in SK-Mel-37 lysates. Other mAbs, however, identified extra bands of an apparent 50-kDa MW, suggesting that they might recognize target epitopes shared with additional proteins and, possibly, with other members of the MAGE-A family. Figure 1b reports representative examples of the reactivities observed.

To obtain additional evidence of the specificity of the reagents under investigation, SK-Mel-37, RE, WM115, MZ-2, A375 and Na8 melanoma cell lines were tested for MAGE-A10 expression by qRT-PCR. Expectably, all lines expressed β-actin housekeeping gene, whereas MAGE-A10 was expressed to decreasing extents in RE, MZ-2, SK-Mel, WM115 and A375 but not in Na8 cell line (Fig. 1c). Lysates from all cell lines, equalized in total protein content, were then tested in "Western" blot assays in the presence of the mAbs under investigation. Recognition of the 70-kDa band in cell lysates closely reflected MAGE-A10 gene expression, as detected in the corresponding RNA preparations. A representative blot is reported in Figure 1d.

Specificity assessment and epitope mapping

TAA s of the MAGE family are characterized by high sequence homology. As a result, mAbs raised by using one TAA as immunogen are frequently characterized by extensive
To unambiguously assess the specificity of the mAbs under investigation, we tested them in “Western” blot assays, by using as target recombinant proteins, lysates of *E. coli* cultures transformed with plasmids encoding different MAGE-A or unrelated proteins, following IPTG induction (Fig. 2). Although 77B mAb only recognized its MAGE-A1 putative target protein, 57B mAb recognized, in addition to its putative MAGE-A3 target, also MAGE-A1 and MAGE-A11. Neither reagent was able to identify MAGE-A10. In contrast, a number of antibodies produced during our study only identified MAGE-A10, as shown, for GA11.1 representative mAb in Figure 2. In no case, evidence of recognition of NY-ESO-1 or thioredoxin (TrxA) could be observed.

Capitalizing on these data, we mapped the specific epitope recognized by GA11.1 IgG1 mAb by using a random peptide library. Clones from six single colonies expanded following repeated panning on the mAb under investigation were sequenced (Table 1). Notably, all of them expressed the (FS)YPE motif, detectable at the COOH end of the MAGE-A10 protein. Most importantly, this motif is not present in a number of additional members of the MAGE family, including MAGE-B1,-B2,-B6 and MAGE-A1,-2,-3,-4,-5,-6,-9 and -12.

**MAGE-A10 is a nuclear protein**

Considering its unambiguous specificity, supported by epitope mapping data, we used GA11.1 mAb to identify the intracellular location of MAGE-A10 TAA. Testis sections stained with this mAb showed that the target antigen in healthy testis is exquisitely expressed in spermatogonia and spermatocyte nuclei (Fig. 3), but, at difference with other C/T TAA, e.g., NY-ESO-1,19,20 it is largely undetectable in cell cytoplasms. No other healthy tissue scored positive upon GA11.1 mAb staining (see below).

**MAGE-A10 protein expression in cancers**

Expression of MAGE-A10 gene has been detected in different types of cancer.8,21–29 However, because of the lack of specific reagents, no comprehensive analysis at the protein level on large tumor databases was possible. We used the anti-MAGE-A10 mAbs described above to stain a multitumor TMA, including cancers, benign tumors and corresponding normal tissues for a total of 2,587 samples (Table 2). Expression in benign tumors was extremely rare (*n* = 6/518, 1.1% of cases). Most interestingly, it was detectable in two cases of moderately dysplastic adenomas of the colon and in benign tumors with remarkable malignant transformation potential, such as paraganglioma (n = 2) and a mixed tumor of the salivary glands (*n* = 1).

In contrast, positive staining, limited to tumor cell nuclei, was detectable in 215 of 1,851 cancers (11.6%). For 43 different malignancies, representative numbers of cases (*n* = 20) were available in the TMA under investigation. In 18 of these
Figure 3. Detection of MAGE-A10 cancer/testis tumor-associated antigens in normal testis and in malignancies of different histological origin. A multitumor tissue microarray from paraffin-embedded sections was stained with GA11.1 hybridoma supernatant according to standard methods. Representative examples of positive stainings of different extents are shown. They include: (a) healthy testis, (b, c) urothelial carcinomas, (d, e) larynx carcinomas, (f, g) colon adenocarcinomas, (h, i) serous carcinomas of the endometrium, (j, k) squamous cell lung carcinoma, (l, m) ovarian serous carcinomas and (n, o) seminomas.
cancer types, MAGE-A10-specific staining was detectable in >10% of cases, with average percentages of positive malignant cells ranging between 20 (mesothelioma and seminoma) and >60% (lung and bladder cancers, laryngeal squamous cell carcinoma).

In particular, among gynecological cancers, breast tumors of the medullary subtype and endometrioid adenocarcinomas displayed similar levels of positivity (>10% of cases, with >40% positive tumor cells). Ovarian and endometrium serous adenocarcinomas expressed MAGE-A10 more frequently, as

Figure 3. Continued
Table 2. Distribution of MAGE-A10-specific staining (GA11.1 mAb) across organs and tissue types \((n = 2,587)\)

<table>
<thead>
<tr>
<th>Organ</th>
<th>Tissue group</th>
<th>Total number of cases</th>
<th>Negative expression (&lt;5%)</th>
<th>Positive expression (&gt;5%)</th>
<th>Average percentage of positive cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adrenal gland</td>
<td>Cortical adenoma</td>
<td>12</td>
<td>12 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pheochromocytoma</td>
<td>26</td>
<td>26 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brain</td>
<td>Meningioma</td>
<td>35</td>
<td>35 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glioblastoma multiforme</td>
<td>29</td>
<td>29 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oligodendroglioma</td>
<td>13</td>
<td>13 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>7</td>
<td>7 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td>Ductal cancer</td>
<td>25</td>
<td>25 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lobular cancer</td>
<td>29</td>
<td>29 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medullary cancer</td>
<td>49</td>
<td>44 (89.8)</td>
<td>5 (10.2)</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Tubular cancer</td>
<td>13</td>
<td>13 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mucinous cancer</td>
<td>15</td>
<td>15 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>3</td>
<td>3 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td>Adenoma, mild dysplasia</td>
<td>27</td>
<td>27 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adenoma, moderate dysplasia</td>
<td>51</td>
<td>49 (96.1)</td>
<td>2 (3.9)</td>
<td>22.5</td>
</tr>
<tr>
<td></td>
<td>Adenoma, severe dysplasia</td>
<td>23</td>
<td>23 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adenocarcinoma</td>
<td>46</td>
<td>42 (91.3)</td>
<td>4 (8.7)</td>
<td>38.8</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>1</td>
<td>1 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endometrium</td>
<td>Endometrioid adenocarcinoma</td>
<td>49</td>
<td>44 (89.8)</td>
<td>5 (10.2)</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Serous adenocarcinoma</td>
<td>32</td>
<td>22 (68.8)</td>
<td>10 (31.3)</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>13</td>
<td>13 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Esophagus</td>
<td>Adenocarcinoma</td>
<td>6</td>
<td>6 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Squamous cell carcinoma</td>
<td>31</td>
<td>23 (74.2)</td>
<td>8 (25.8)</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Small cell carcinoma</td>
<td>1</td>
<td>1 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>7</td>
<td>7 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gall bladder</td>
<td>Adenocarcinoma</td>
<td>34</td>
<td>29 (85.3)</td>
<td>5 (14.7)</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>3</td>
<td>3 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kidney</td>
<td>Clear cell renal cell carcinoma</td>
<td>50</td>
<td>50 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Papillary renal cell carcinoma</td>
<td>25</td>
<td>25 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chromophobe renal cell carcinoma</td>
<td>10</td>
<td>9 (90.0)</td>
<td>1 (10.0)</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Oncocytoma</td>
<td>19</td>
<td>19 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>16</td>
<td>16 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larynx</td>
<td>Squamous cell carcinoma</td>
<td>24</td>
<td>19 (79.2)</td>
<td>5 (20.8)</td>
<td>61</td>
</tr>
<tr>
<td>Liver</td>
<td>Hepatocellular carcinoma</td>
<td>82</td>
<td>78 (95.1)</td>
<td>4 (4.9)</td>
<td>13.8</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>15</td>
<td>15 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lung</td>
<td>Squamous cell carcinoma</td>
<td>43</td>
<td>28 (65.1)</td>
<td>15 (34.9)</td>
<td>65.3</td>
</tr>
<tr>
<td></td>
<td>Adenocarcinoma</td>
<td>75</td>
<td>66 (88.0)</td>
<td>9 (12.0)</td>
<td>56.1</td>
</tr>
<tr>
<td></td>
<td>Large cell carcinoma</td>
<td>20</td>
<td>17 (85.0)</td>
<td>3 (15.0)</td>
<td>63.3</td>
</tr>
<tr>
<td></td>
<td>Small cell carcinoma</td>
<td>44</td>
<td>38 (86.4)</td>
<td>6 (13.4)</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Bronchioloalveolar adenocarcinoma</td>
<td>6</td>
<td>6 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>11</td>
<td>11 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymphatic tissue</td>
<td>Diffuse large B-cell lymphoma</td>
<td>10</td>
<td>10 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-Hodgkin lymphoma, others</td>
<td>52</td>
<td>52 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mixed cellularity Hodgkin lymphoma</td>
<td>14</td>
<td>14 (100)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2. Distribution of MAGE-A10-specific staining (GA11.1 mAb) across organs and tissue types ($n = 2,587$) (Continued)

<table>
<thead>
<tr>
<th>Organ</th>
<th>Tissue group</th>
<th>Total number of cases</th>
<th>Negative expression ($&lt;5%$)</th>
<th>Positive expression ($&gt;5%$)</th>
<th>Average percentage of positive cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodular sclerosis Hodgkin lymphoma</td>
<td>23</td>
<td>22 (95.7)</td>
<td>1 (4.4)</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>13</td>
<td>13 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myometrium</td>
<td>Leiomyoma</td>
<td>47</td>
<td>47 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>9</td>
<td>9 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroendocrine tissue</td>
<td>Extra-adrenal paraganglioma</td>
<td>8</td>
<td>6 (75.0)</td>
<td>2 (25.0)</td>
<td>15</td>
</tr>
<tr>
<td>Typical carcinoid of the lung</td>
<td>33</td>
<td>33 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral cavity</td>
<td>Squamous cell carcinoma</td>
<td>42</td>
<td>38 (90.5)</td>
<td>4 (9.5)</td>
<td>46.3</td>
</tr>
<tr>
<td>Normal</td>
<td>13</td>
<td>13 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroendocrine tissue</td>
<td>Serous adenocarcinoma</td>
<td>40</td>
<td>32 (80.0)</td>
<td>8 (20.0)</td>
<td>36.3</td>
</tr>
<tr>
<td>Mucinous adenocarcinoma</td>
<td>15</td>
<td>15 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral cavity</td>
<td>Endometrioid adenocarcinoma</td>
<td>42</td>
<td>41 (97.6)</td>
<td>1 (2.4)</td>
<td>25</td>
</tr>
<tr>
<td>Normal</td>
<td>9</td>
<td>9 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parathyroid</td>
<td>Adenoma</td>
<td>31</td>
<td>31 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>2</td>
<td>2 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peripheral nerves</td>
<td>Neurofibroma</td>
<td>17</td>
<td>17 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schwannoma</td>
<td>23</td>
<td>23 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pleura</td>
<td>Mesothelioma</td>
<td>28</td>
<td>23 (82.1)</td>
<td>5 (17.9)</td>
<td>21</td>
</tr>
<tr>
<td>Prostate</td>
<td>Adenocarcinoma, hormone-refractory</td>
<td>29</td>
<td>29 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma, untreated</td>
<td>48</td>
<td>48 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>17</td>
<td>17 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saliory gland</td>
<td>Warthin tumor</td>
<td>18</td>
<td>18 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pleomorphic adenoma</td>
<td>39</td>
<td>38 (97.4)</td>
<td>1 (2.6)</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>Adenoid cystic carcinoma</td>
<td>24</td>
<td>24 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>13</td>
<td>13 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skin</td>
<td>Basalioma</td>
<td>59</td>
<td>40 (67.8)</td>
<td>19 (32.2)</td>
<td>50.5</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>33</td>
<td>31 (93.9)</td>
<td>2 (6.1)</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Appendageal tumors (benign)</td>
<td>11</td>
<td>11 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malignant melanoma</td>
<td>50</td>
<td>31 (62.0)</td>
<td>19 (38.0)</td>
<td>45.8</td>
<td></td>
</tr>
<tr>
<td>Benign nevus</td>
<td>12</td>
<td>12 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibrous histiocytoma</td>
<td>12</td>
<td>11 (91.7)</td>
<td>1 (8.3)</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Kapillary hemangioma</td>
<td>22</td>
<td>22 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaposi sarcoma</td>
<td>15</td>
<td>15 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>6</td>
<td>6 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small intestine</td>
<td>Adenocarcinoma</td>
<td>22</td>
<td>22 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>4</td>
<td>4 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soft tissue</td>
<td>Liposarcoma</td>
<td>22</td>
<td>22 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malignant fibrous histiocytoma</td>
<td>26</td>
<td>24 (92.3)</td>
<td>2 (7.8)</td>
<td>62.5</td>
<td></td>
</tr>
<tr>
<td>Leiomysarcoma</td>
<td>45</td>
<td>43 (95.6)</td>
<td>2 (4.4)</td>
<td>52.5</td>
<td></td>
</tr>
<tr>
<td>Tendon sheath, giant cell tumor</td>
<td>14</td>
<td>14 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal (skeletal muscle)</td>
<td>16</td>
<td>16 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal (smooth muscle)</td>
<td>5</td>
<td>5 (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
detected by immunostaining (30% and 20% of cases, with >50 and >30% positive tumor cells, respectively). MAGE-A10-specific staining was detectable in squamous cell carcinomas of the esophagus (8 of 31 cases in 25% of cells on an average) and of the larynx (5 of 24 cases in >60% of tumor cells on an average).

Notably, MAGE-A10 protein expression appears to be particularly frequent in lung cancers. In more than 34% of squamous cell carcinomas and 12%, 15%, and 13% of adenocarcinomas, large, and small cell carcinomas, respectively, more than 50% of tumor cells do express MAGE-A10 protein. Positivities were also observed in mesothelioma (5 of 28 cases), albeit with relatively low average numbers of stained tumor cells (20%).

Skin malignancies frequently expressed MAGE-A10. In particular, in 19 of 59 basalioma, more than 50% of tumor cells expressed the C/T TAA under investigation. Similarly, in 19 of 50 melanoma (38%), on an average, 45% of tumor cells stained positive with anti-MAGE-A10 GA11.1 mAb. Positivities were also observed in mesothelioma (5 of 28 cases), albeit with relatively low average numbers of stained tumor cells (20%).

Skin malignancies frequently expressed MAGE-A10. In particular, in 19 of 59 basalioma, more than 50% of tumor cells expressed the C/T TAA under investigation. Similarly, in 19 of 50 melanoma (38%), on an average, 45% of tumor cells stained positive with anti-MAGE-A10 GA11.1 mAb. Positivities were also observed in mesothelioma (5 of 28 cases), albeit with relatively low average numbers of stained tumor cells (20%).

Skin malignancies frequently expressed MAGE-A10. In particular, in 19 of 59 basalioma, more than 50% of tumor cells expressed the C/T TAA under investigation. Similarly, in 19 of 50 melanoma (38%), on an average, 45% of tumor cells stained positive with anti-MAGE-A10 GA11.1 mAb. Positivities were also observed in mesothelioma (5 of 28 cases), albeit with relatively low average numbers of stained tumor cells (20%).

Skin malignancies frequently expressed MAGE-A10. In particular, in 19 of 59 basalioma, more than 50% of tumor cells expressed the C/T TAA under investigation. Similarly, in 19 of 50 melanoma (38%), on an average, 45% of tumor cells stained positive with anti-MAGE-A10 GA11.1 mAb. Positivities were also observed in mesothelioma (5 of 28 cases), albeit with relatively low average numbers of stained tumor cells (20%).

Skin malignancies frequently expressed MAGE-A10. In particular, in 19 of 59 basalioma, more than 50% of tumor cells expressed the C/T TAA under investigation. Similarly, in 19 of 50 melanoma (38%), on an average, 45% of tumor cells stained positive with anti-MAGE-A10 GA11.1 mAb. Positivities were also observed in mesothelioma (5 of 28 cases), albeit with relatively low average numbers of stained tumor cells (20%).
Here, we report the generation and the characterization of MAGE-A10-specific mAbs. One of these reagents, recognizing a discrete MAGE-A10-specific epitope, has been used to stain a multitissue array comprising more than 2,000 specimens and including malignancies and benign tumors of different histological origin together with corresponding healthy tissues. This reagent has helped to reveal similarities and differences between MAGE-A10 and the other members of the MAGE-A family. First, this TAA is exclusively intranuclear. In this context, its relatively high immunogenicity might be related to the release of nuclear proteins following necrotic cell death, detectable in specific areas, in different types of tumors. Second, it identifies relatively large percentages of tumor cells expressing the target proteins in cancers known to express C/T TAA, including squamous cell carcinoma of different histological origin, melanoma, lung and gynecological malignancies and bladder cancers. Third, MAGE-A10 protein, at difference with other C/T TAA, appears to be of limited expression in soft tissue cancers. Fourth, in contrast, it appears to be relatively frequently detectable in gall bladder tumors and, interestingly, in a sizeable subgroup of colorectal cancers.

Considering the high potential immunogenicity of MAGE-A10, the reagents described in our study might prove of critical relevance in the development of targeted active, specific immunotherapies by helping to establish clear indications for treatment and in the monitoring of their impact on tumor progression. Furthermore, as the prognostic relevance of C/T TAA expression in tumors of diverse histological origin is emerging, the availability of these mAbs might set the stage for investigations addressing the role of MAGE-A10 expression in individual tumor entities.

On the other hand, the biological function of the proteins of the MAGE-A family remains largely elusive. MAGE-A gene expression has been shown to suppress p53-dependent apoptosis and to promote the “in vitro” and “in vivo” viability of mast cell lines. Previous studies in a thyroid carcinoma model have suggested that selected MAGE-A genes might control fibronectin-mediated tumor progression. Notably, MAGE-A11, also characterized by a putatively specific intranuclear location, has been shown to coregulate androgen receptor-mediated transcriptional activity.

Most interestingly, MAGE-A gene expression has been found to be correlated with genome-wide demethylation, frequently representing an early event during carcinogenesis, associated with hypermethylation of defined tumor suppressor genes. Strikingly, a number of tumors showing evidence of high MAGE-A10 expression in our study, including lung, skin, bladder and head and neck cancers, are also known to be characterized by a frequent genome-wide hypomethylation. Also, considering the specific nuclear location of MAGE-A10, further studies are warranted to comparatively address its expression together with the DNA methylation status of repetitive elements, including long interspersed nuclear elements (LINE-1).

Acknowledgements

Our study was partly supported by SNF grants to G.C.S. and L.Te.

16. Gure AO, Stockert E, Arden KC, Boyer AD, Viars CS, Scanlan MJ, Old LJ, Chen YT. CT10: a new cancer-testis (CT) antigen homologous to CT7 and the MAGE family, identified by representational-difference analysis. 


7. GENERAL DISCUSSION

7.1 3’UTR poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair deficient cancers

In this study we describe a mononucleotide (T/U)\textsubscript{16} tract, EWS16T, located in the 3’ UTR of the Ewing sarcoma break point region 1 (EWSR1) gene which discriminates MMR proficient from MMR deficient cancers with 100% sensitivity and 99.5% specificity. We demonstrate \textit{in vitro} and \textit{in vivo} that contractions at this locus alter poly(A) site selection by promoting SFPQ-mediated distal poly(A) site usage in EWSR1 pre-mRNAs and result in decreased mRNA as well as protein expression. In contrast to their proficient counterparts, MMR deficient CRC display altered subcellular localization of EWS with diffuse cytoplasmic staining. EWS16T thus not only represents a novel monomorphic MSI target locus to accurately identify both, hereditary and sporadic, MMR deficient cancers but contractions therein affect multiple regulatory mechanisms implicating the RNA-/DNA-binding protein EWS in MSI-associated colorectal tumorigenesis.

The Ewing sarcoma (EWS) protein is a member of the TET family (TLS/FUS, EWS, and TAF15) of RNA- and DNA-binding proteins, with proposed functions in transcription and RNA processing. The domain composition of TET proteins includes a transcription activation domain at the N terminus and RNA-binding domains, including three RGG boxes and one RRM motif, at the C terminus. Additional domains harbored by TET proteins include an IQ domain, which interacts with calmodulin and is phosphorylated by PKC [198], and one zinc finger motif [199]. EWS interacts with the preinitiation complex TFIID and with subunits of the RNA polymerase II (RNAPII) [200], suggesting its involvement in transcriptional regulation. EWS also interacts with splicing factors, including the U1 snRNP protein U1C, which recognizes 5’ splice sites [201], the branchpoint binding protein BBP/SF1 [202], and the spliceosome component YB-1 [203, 204], suggesting a function for EWS in pre-mRNA splicing. Consistent with this potential dual role, EWS has been shown to regulate cyclin D1 transcripts both transcriptionally and at the level of splicing, with
the oncogenic fusion protein EWS-FLI1 promoting the expression of the oncogenic cyclin D1b splice variant in Ewing sarcoma cells [205]. More recently, EWS has been shown to regulate alternative splicing (AS) of the p53 repressor MDM2 [204, 206]. In addition EWS has been described as component of the microprocessor complex that mediates the genesis of microRNAs [207].

The physiological role of EWS is largely unknown but based on its structural properties this protein is thought to be involved in diverse processes including gene expression, RNA processing / transport and cell signaling. Knockout of EWS in mice results in postnatal lethality, defects in pre-B cell development, meiotic arrest/germ cell apoptosis, premature cellular senescence, and hypersensitivity to ionizing radiation (IR) [208]. These observations suggest additional roles for EWS in homologous recombination, DNA damage response, and maintenance of genome integrity [206].

With respect to tumorigenesis, genetic alterations in EWSR1 were first observed in Ewing sarcoma, the second most common malignant bone tumor in children [209, 210]. The EWSR1-FLI1 fusion is the most common, being found in 85% of the cases [211]. The fusion protein retains the N-terminal transcription activation domain but loses the RNA-binding domains, which are replaced with the DNA binding domain of the fusion partner. The fusion proteins are constitutively active and have been shown to alter the transcription of several downstream targets. Ewing Sarcoma is largely thought of as a gain of function phenotype. Loss of the normal EWSR1 function has been largely overlooked, in spite of the fact that the protein has a very canonical RNA binding domain and has been shown to regulate several RNA processing events in the nucleus [212, 213].

Our results described in details in chapter I suggest that the poly T/U tract in 3’UTR of EWSR1 gene (EWS16T) represents a novel, quasi-monomorphic MSI target locus which identifies both, hereditary and sporadic, MMR deficient cancers with 100% sensitivity and 99.5% specificity. The contractions at this locus affect multiple regulatory mechanisms including alternative polyadenylation, mRNA / protein expression and possibly subcellular localization thereby implicating the RNA-/DNA-binding protein EWS, in MSI-associated colorectal tumorigenesis. Furthermore we demonstrated the biological effects of MSI-associated 3’UTR contractions on gene expression in vivo and in vitro for the first time.

In summary, due to only few and sketchy data concerning the role of EWS in normal
cell physiology and in CRC-related tumorigenesis, future investigations are needed. In particular to comprehensively characterize the major EWSR1–related downstream targets/pathways involved in cell physiology and differentiation the transcriptomes and miRNAomes of human fibroblasts and their progenitor cells, the mesenchymal stem cells should be investigated. On the other hand, The role(s) of EWSR1 in colorectal carcinogenesis by the generation of stable EWSR1 overexpressing/downregulating normal and CRC cell lines, in order to investigate EWSR1 effect on apoptosis and cell migration in vitro as well as its impact in tumorigenesis and it metastatic potential in vitro and in vivo.
It is well accepted that the key genetic mutations underlying initiation, progression and transformation of adenomas into CRC include *APC*, *KRAS* and *TP53*, respectively [214]. Mutations of known genes are identified in 60%, 35% and 50% [215] suggesting the existence of more than just one single dominant pathway to CRC development [216].

Heterozygous loss of variable parts of chromosome 8p constitutes a frequent feature of CRC [217] and has been linked to DNA breakage at fragile sites located at 8p12 and 8p22 [218]. Tumour suppressor genes have not been localized to this critical chromosomal region, thus precluding the identification of a likely mechanism for how genetic alterations in 8p contribute to CRC development and progression.

Looking at Oncomine signature of mRNA, we noted that SH2D4A protein was deregulated in various cancer types but no information concerning colorectal cancer were available. We started to investigate LOH and copy number variation in an unselected cohort of 70 CRC patients. In 27 subjects the two *SH2D4A* alleles could be separated by microsatellite markers and SNPs. In 14 of these patients (52%), the primary tumour had lost or diminished *SH2D4A* expression. The rate of metastasis was significantly increased among these tumours (12/14; 86%) when compared to *SH2D4A* expressing primary CRC (3/13). Using 3 microsatellite markers and 5 SNPs, SH2D4A LOH was detected in 7 of the 14 tumours (50%) marked by a partial or complete lack of *SH2D4A* expression. Gene dosage quantification of the short arm of chromosome 8 revealed a monoallelic deletion in 6 and a biallelic deletion in one of these 7 tumours with LOH. Though 4 of the 6 patients with monoallelic tumours were heterozygous for the intronic SNP rs17128221 (c.342-5T>C) the T allele was selectively lost in their tumour and the C allele provided a splice donor site causing the skipping of exon 4 and a premature termination of translation in exon 5.

In addition we demonstrated that *SH2D4A*, physically interacts with the EGFR/STAT3 pathway and controls cell proliferation. Upon EGF signalling, *SH2D4A* protein recruits the serine/threonine phosphatase PP1β to the receptor complex and represses activated STAT3 via dephosphorylation. *SH2D4A* expression reduces anchorage-independent tumour cell growth and its loss promotes the expression of c-Myc, Cyclin D1 and Jun B.
Recently in accordance with our data Roessler et al. published unsupervised analyses of array comparative genomic hybridization data associated loss of chromosome 8p with poor outcome (reduced survival); somatic copy number alterations correlated with expression of 27.3% of genes analyzed. They associated expression levels of 10 of these genes with patient survival in 2 independent cohorts (comprising 319 cases of hepatocellular carcinoma (HCC) with mixed etiology) and 3 breast cancer cohorts (637 cases). Among the 10-gene signature, a cluster of 6 genes on 8p, (DLC1, CCDC25, ELP3, PROSC, SH2D4A, and SORBS3) were deleted in HCCs from patients with poor outcomes. In vitro and in vivo analyses indicated that the products of PROSC, SH2D4A, and SORBS3 have tumor-suppressive activities, along with the known tumor suppressor gene DLC1 [219].

Thus, SH2D4A could represent novel tumor suppressor gene acting in different tumor entities. Due to its interaction with STAT3 in the control of the EGFR signaling pathway, which is involved in the development and progression of several human tumors, including colorectal cancer [220], it may be represent a novel promising target for CRC treatment.

In order to understand if germline mutations in SH2D4A could contribute in familial colorectal cancer development, whole genome sequencing of 200 index patients from familial colorectal cancer patients, without mutations in any of the known genes, will be done.
7.3 HMGA proteins as prognostic markers in different tumor entities

Our work demonstrates that the percentage of tumour cells showing HMGA1 and HMGA2 nuclear immunoreactivity correlates positively with increasing malignancy of breast and pancreatic tumors. These studies indicate that the overexpression of HMGA1 and HMGA2 is linked to proliferation in the cancer and may have important implications, promoting the growth and spread of tumors. On the molecular basis HMGA1 and HMGA2 interact with several transcription factors, influence gene expression patterns and regulate cell growth, differentiation, apoptosis, and transformation [156]. Their expression has been detected in many kinds of benign and malignant tumors and it is associated with a highly malignant phenotype (poor prognostic index) [221].

HMGA is high wexpressed during embryogenesis, but is undetectable or very low in differentiated adult tissues [222, 223], being confined, at least for HMGA2, to the staminal compartment [224-226]. In vivo studies revealed an important role of HMGA proteins in adipogenesis [222, 225], somatic growth [227], cardiac cell growth control [228] and glucose homeostasis [229, 230]. In accordance with this, HMGA mutations have been detected in human diseases such as lipomas [231-233], gigantism [234], dwarfism [235], and diabetes [229]. Moreover, HMGA2 has also been suggested recently by genome-wide SNP studies to influence human height variation [236, 237]. HMGA overexpression is a constant feature of human malignant neoplasms and is frequently associated, with or without gene rearrangements, with human tumors [221, 238]. In addition HMGA proteins also regulate the transcription of genes that are involved in DNA repair. Reeves et al. described a number of genes involved in DNA repair that were negatively regulated by HMGA1 in MCF7 human breast-cancer cells, suggesting that HMGA proteins can influence DNA repair by negatively regulating the transcriptional activity of genes involved in various aspects of DNA-damage recognition and removal [239]. Consistently, in another study made by Borrmann et al. HMGA2 has been linked to the promoter of the nucleotide excision-repair gene ERCC1 where it negatively modulates its activity [240]. Moreover, it has been shown that HMGA1 can downregulate BRCA1 expression, which is involved in homologous recombination, by binding directly to its promoter region, and that there
is an inverse correlation between HMGA1 and BRCA1 expression in human breast carcinomas [241]. Consistent with the role of BRCA1 in DNA double-strand break (DSB) repair, it has been shown that HMGA proteins potentiate genotoxic stress induced by different DNA-damaging agents causing DSBs, such as cisplatin, bleomycin, doxorubicin and X-rays [242, 243]. HMGA proteins can indirectly inhibit DNA repair through cyclin A induction, and it recently has been reported that the cyclin A1–cyclin-dependent kinase 2 complex also regulates DSB repair [221, 244]. Essentially, the data reported in this thesis indicate that HMGA1 and HMG2 are implicated in carcinogenesis and may play a role in the development of a particular phenotype of breast and pancreas carcinomas. Therefore, HMGA1 and HMG2 expression may represent an indicator of poor prognosis of human tested cancers. Then studies are needed to eluate their role as prognostic factor in tumor progression and maybe as potential target in molecular therapy.

To understand which are the downstream targets directly regulated by HMGA1 and HMGA2, a transcriptome analysis of different tumors developed in HMGA1 KO mice (Prof. Alfredo Fusco laboratory) is planned; this strategy allow us to define an expression pathway and thus understand the relationship with the most important tumour related pathway and test them in vivo.
7.4 **MAGE-A10 overexpression in lung, skin and urothelial malignancies**

In this study we produced a MAGE-A10 protein in soluble recombinant form, and used it to generate specific monoclonal antibodies (mAbs). One of these reagents, recognizing an epitope located at the COOH terminus of the MAGE-A10 gene product, was used to stain a multitumor tissue microarray comprising more than 2,500 paraffin-embedded specimens including healthy tissues, benign tumors and malignancies of different histological origin. MAGE-A10 was identified as an intranuclear protein of an apparent molecular weight of 70 kDa, expressed in normal spermatogonia and spermatocytes but in no other healthy tissue. Most importantly, this cancer/testis tumor associated antigen (C/T TAA) appears to be expressed in high (>50%) percentages of cancer cells from a number of malignancies, including lung, skin and urothelial tumors. Unexpectedly, high expression of MAGE-A10 TAA at the protein level was also detectable in gynecological malignancies and stomach and gall bladder cancers.

MAGE-A proteins are known to be highly expressed in a wide range of cancers like breast, ovary, lung, skin, urothelial and bladder [181, 245-248] and their expression is observed mainly in cancers with malignant phenotypes, [181]. MAGE detection also correlates with poor prognosis in cancer patients, underpinning the idea that MAGE proteins may contribute actively towards malignancy [247]. The mechanism by which MAGE-A overexpression occurs in tumor cells is not totally understood. Usually the expression of MAGE-A in somatic tissues is repressed by DNA hypermethylation of CpG dinucleotides in promoters, which acts to prevent access of transcription factors like Ets and SP1 [249, 250]. In tumor cells epigenetic reprogramming can result in promoter hypo-methylation leading to the aberrant expression of one or more of these genes, in fact MAGE-A expression can be induced by demethylating agents such as 5-aza-2'-deoxycytidine in non-expressing cells of various origins [249, 250]. Seems to be also that chromatin remodelling events which occur during tumour development, like histone acetylation and methylation, can contribute to changes in MAGE-A levels in cancer cells and, at least in some circumstances, can be under control of hormones such as FGF, estrogen, leutenising hormone and directly regulated by microRNAs miR-34a [247, 251-254]. Tumors also show significant differences in signal localization (nuclear vs.
cytoplasmic) [245], suggesting that either different MAGE-A family members display differential localization or that the mechanisms controlling localization are cancer cell-specific.

Several studies have established that MAGE-A proteins can repress p53-mediated transcription, through direct and indirect mechanisms, and inhibit both p53-mediated apoptosis and senescence, two major tumour suppressor mechanisms utilized by p53 [247, 255-259]. These studies also establish the principle that elimination of MAGE-A expression in cultured cells is sufficient to induce p53-mediated apoptosis without the need to use genotoxic agents. This suggests that the development of compounds that block the p53/MAGE-A interaction in cancer cells expressing these proteins could have enormous therapeutic potential that minimizes the requirement for genotoxic approaches and the accompanying side effects. Advances in our knowledge of MAGE structure and interaction with partner proteins are beginning to cover the way towards developing such therapeutic approaches.
8. APPENDIX

8.1 microRNA expression profiling in mismatch repair associated colorectal cancer

My contribution to this work:
• Assessment and optimization of miRNA extraction techniques;
• Total RNA and microRNA extraction from CRC cell lines and human samples;
• miRNA library generation;
• Analysis of deep sequencing data;
• Statistical comparison;
microRNA expression profiling in mismatch repair associated colorectal cancer

Salvatore Piscuoglio¹, Shivendra Kishore², Mihaela Zavolan¹, Luigi Terracciano³ and Karl Heinimann¹

¹ Research Group Human Genetics, Department of Biomedicine, University of Basel, and Division of Medical Genetics, University Children's Hospital, Basel, Switzerland.
² Biozentrum, University of Basel, Basel, Switzerland.
³ Institute of Pathology, University Hospital of Basel, Basel, Switzerland
Abstract

Lynch syndrome represents the most common, autosomal dominantly inherited cancer predisposition worldwide and is characterised by early onset colorectal cancer around age 44 years. The syndrome accounts for 3 to 5% of all colorectal cancers in Switzerland and is caused by germ line mutations in DNA mismatch repair (MMR) genes, predominantly MLH1 and MSH2. MMR deficiency leads to genomic instability in the tumor, resulting in the genome-wide accumulation of somatic mutations, in particular at short repetitive sequences giving rise to microsatellite instability (MSI). To date, little is known on the specific molecular genetic alterations, which initiate and promote cancer development in Lynch syndrome patients.

microRNAs (miRNAs) are a family of small non-coding RNAs which are thought to control gene expression of about 30% of all protein-coding genes in humans. They regulate a variety of cellular processes and are likely to have a causal role in carcinogenesis since they are altered in most cancer types, including colon cancer. Their diagnostic, prognostic and therapeutic potential has fuelled miRNA research in recent years. Despite advances on the role of MSI status and miRNA expression in sporadic colorectal cancer, a comprehensive study on miRNA profiles in Lynch syndrome-associated colorectal cancers has not been performed to date.

In a long, painstaking process we evaluated and scrutinized several extraction methods and protocols to obtain high-quality miRNA from snap-frozen MMR proficient and deficient cancers as well as cancer cell lines. Finally, radiolabelling of total RNA with $^{32}$P followed by direct excision of the miRNA fraction from a 8-15% polyacrylamide gel and extraction with TRI reagent solution proved to yield the best miRNA quality. We therefore applied the Truseq sequencing on a HiSeq2000 platform and investigated the miRNA profiles by unsupervised analysis using hierarchical clustering and principal component analysis generating a initial list of top/bottom interesting miRNAs.

Currently, selected top most differentially expressed miRNAs will be verified / validated by quantitative real-time PCR on a total of 100 Lynch syndrome and 50 sporadic colorectal cancers. Subsequently, tissue expression of verified miRNAs and potential target proteins will be assessed by in situ hybridization and immunohistochemistry, respectively. Finally, selected miRNAs will be functionally characterised by cell transfection experiments.
Introduction

About 15-20% of all colorectal cancers (CRCs) are thought to arise from an inherited genetic susceptibility to colorectal adenomas and carcinomas [1]. Lynch syndrome (formerly known as hereditary non-polyposis colorectal cancer, HNPCC), is an autosomal dominantly inherited cancer condition with an estimated carrier frequency of 1:2000 (up to 1:200) and thus represents the most common inherited cancer predisposition worldwide. It is estimated to account for about 2-5% of the total colorectal (CRC) cancer burden [2].

It is characterized by the development of colorectal as well as a distinct spectrum of extracolonic cancers (predominantly of the endometrium, ovary, stomach), usually diagnosed before age 50 years. The syndrome is caused by germ line mutations in the mismatch repair (MMR) genes MLH1, MSH2, MSH6, and PMS2, which recognize and correct errors that occur during DNA replication. MSH2 and MLH1 are mutated in about 80% to 90% of patients with typical Lynch syndrome whereas germ line mutations in MSH6 and PMS2 tend to result in a less severe phenotype [3]. Following somatic inactivation of the wild-type allele in the cancer cell MMR deficiency leads to genome-wide accumulation of replication errors predominantly at microsatellite loci, the hallmark lesion of Lynch syndrome, termed microsatellite instability (MSI). In the clinical setting, immunohistochemical assessment of the presence / absence of the MMR proteins in the tumor is used as a preliminary tool, often in conjunction with microsatellite instability testing, to assess MMR proficiency/deficiency and to determine who may benefit from germ line testing [4].

microRNAs (miRNAs) represent a large family of small non-coding RNAs of about 21 nucleotides in length that serve as effector molecules of sequence-specific gene silencing [5]. It is estimated that the number of miRNAs in the human genome range from about 450 to 1000 and that they control gene expression of about 30% of all protein-coding genes in mammals. The majority of identified miRNAs, currently estimated are highly evolutionary conserved among many distantly related species, from worms to human, suggesting that miRNAs have very important roles in essential biological processes, including developmental timing, stem-cell differentiation, signal transduction, disease and cancer. Their causal role in carcinogenesis is further substantiated by the fact that miRNAs are altered in most cancer types, including colon cancer [6].
In the nucleus, miRNA genes are generally transcribed by RNA polymerase II or III to form large primary miRNA transcripts (pri-miRNAs). These are further processed by Drosha, a RNase III protein, into 70-nucleotide miRNA precursors (pre-miRNA). After transport into the cytoplasm, pre-miRNAs are further processed by another RNase III enzyme, Dicer, into miRNA duplexes, typically consisting of 19-25 nucleotides in length [7, 8]. Subsequently these duplexes can be loaded into the miRNA-associated multiprotein RNA-induced silencing complex (miRISC) and the mature miRNA strand is preferentially retained. Once bound to the 3’UTR of target mRNAs, the mature miRNA induces cleavage, translational repression or deadenylation, depending on the degree of complementarity [9]. A single miRNA may bind to as many as 200 target genes encoding a broad range of proteins, such as transcription factors, receptors and transporters. In recent years, several approaches have been used to identify miRNA targets [10].

With the advent of massively parallel (“next generation”) sequencing technology genome-wide miRNA profiling has now become possible allowing global assessment of expression-regulating mechanisms in sporadic and hereditary cancer [11, 12]. The results of this study will enhance the scarce knowledge available to date on the role of miRNAs in Lynch syndrome-associated colorectal cancer development. The findings are likely to shed light on intestinal carcinogenesis in general and allow further characterisation of the various pathways involved in MSI-related tumorigenesis. The identification and characterisation of miRNA expression signatures in Lynch syndrome cancer patients may not only improve clinical risk assessment and prognosis, but also offer an opportunity to identify novel therapeutic strategies for cancer patients.
Materials and methods

RNA Isolation
To isolate genomic total RNA from cell lines Tri-Reagent (Ambion) were used according to the manufacturers’ guidelines.

Cell lines
Six colorectal cancer cell lines from the American Type Culture Collection (ATCC, Rockville, MD) were used for this study: four repair deficient cell lines (HCT116, LoVo, HCT15, DLD-1) and two mismatch repair proficient (SW480 and HT29). HCT116, HCT15 and DLD-1 cells (ATCC, Rockville, MD) were cultured in RPMI 1640 (Invitrogen Basel, Switzerland) supplemented with 10% fetal bovine serum FBS, 1% Kanamycin sulphate, 1% GlutaMAX-I, 1% Sodium Pyruvate, 1% non Essential Amino Acids (NEAA), 1% HEPES (all from Invitrogen Basel, Switzerland) and 0.1% 2-mercapto-ethanol (Sigma-Aldrich Basel, Switzerland). HT29 cells were grown in McCoy's 5A Medium (Invitrogen Basel, Switzerland) with 10% fetal bovine serum FBS, Kanamycin sulphate and GlutaMAX-I (all from Invitrogen Basel, Switzerland). SW480 cells were cultured in L-15 Medium (Sigma-Aldrich Basel, Switzerland) with 10% FBS, 1% GlutaMAX-I and 1% Kanamycin sulphate (all from Invitrogen Basel, Switzerland). Cells were maintained at 37°C with 5% CO2.
Preliminary results
To yield the best miRNA quality the total RNA (typically 2 µg) marked radioactively with P\(^{32}\) from each sample was run on denaturing polyacrylamide-urea gels. The approximately 17-25 nucleotide RNAs were excised from the gel, ligated to sequencing adaptors on both ends, and reverse-transcribed according to the manufacturers’ guidelines of Illumina. The resulting cDNA library was PCR-amplified for 15 cycles and gel-purified on 6% acrylamide gel. The gel-purified amplicon quality and quantity were analyzed on a 6% acrylamide gel relative to oligonucleotides of known concentration and size. The library (120 µL 1-4pM) was loaded on an Illumina HiSeq 2000 system (Illumina) on a flow cell according to the manufacturer's instructions, where DNA molecules were attached to high-density universal adaptors in the flow cells and amplified. The DNA clusters generated via this process were sequenced with sequencing-by-synthesis technology, where successive high-resolution images of the 4-color fluorescence excitation dependent on the base incorporated during each cycle were captured [13]. Then, the data were pre-processed, including steps of quality check and normalization. miRNA profiles were investigated by unsupervised analysis using hierarchical clustering and principal component analysis (PCA). For computational prediction of miRNA targets and pathway analysis the MirZ web server was used which provides statistical analysis and data mining tools operating on up-to-date databases of sequencing-based miRNA expression profiles and of predicted miRNA target sites [14].

In total, we identified 1240 miRNAs differentially expressed in MMR deficient compared to proficient cancers cell lines. Among the most downregulated miRNAs (Table 1) were hsa-miR-371-5p, involved in Wnt/beta-catenin signaling, a crucial pathway for colorectal carcinogenesis [15], and hsa-miR-200a, which has been shown to inhibit the epithelial-mesenchymal transition [16] and to be deregulated in bladder breast and endometrial cancer (which are, intriguingly, part of the LS-tumor spectrum) [17, 18]. Among the most upregulated miRNAs we observed hsa-miR-141 which is believed to modulate the oxidative stress response [19] and hsa-miR-125b, a negative regulator of p53, also referred to as the “guardian of the genome”[20].

Currently we are analyzing the deep sequencing data of 4 cancer specimens matched with their normal counterpart from 2 LS-related, microsatellite-unstable asa well as 2 sporadic microsatellite-stable CRCs.
**Brief discussion**

MMR-deficient CRCs exhibit mRNA and miRNA expression profiles distinct from their stable, MMR-proficient counterparts. The study by Di Pietro et al. [21], in which our research group also took part, analysed gene expression in proximal colon cancers and was able to divide them into two groups that almost perfectly corresponded with their MMR status. In addition, expression changes in genes involved in apoptosis and the immune response were consistent with the better prognosis of MMR-deficient cancers. Kruhoffer et al. [22] studied mRNA expression in 34 MSI-High (mostly sporadic) and 67 MSS stage II and III colorectal cancers and devised a gene expression signature based on nine genes able to distinguish MMR status.

To determine the degree of gene expression differences which could be explained by CpG island methylation, we assessed the presence of CpG islands in and 5kb up- and down-stream of the top most differentially expressed genes from both studies using the UCSC genome database: only about 50-60% of these genes harbour CpG islands. Thus, about 40-50% of the gene expression changes in MMR-deficient CRCs cannot readily be explained by epigenetic regulation and are likely to be caused by other mechanisms, in particular deregulation of miRNA expression.

The mutator phenotype that results from MMR dysfunction induces the acquisition of additional gene mutations that promote cancer progression [23]. In addition to germline mutations, various pathogenic events, including promoter methylation [24] and reduced histone acetylation [25], result in reduced or absent expression of core MMR proteins, as do microenvironmental factors, such as inflammation and hypoxia [26]. Our preliminary results suggest that some miRNA such as miR-155, could play a role in this multifactorial regulation by causing down-modulation of the core MMR heterodimeric proteins MSH2-MSH6 and MLH1-PMS2. The simultaneous inhibition of these essential MMR components by some miRNA could well explain the observed mutator phenotype. Thus far, only little is known about the role of MSI status and miRNA expression in CRC. A study by Lanza et al. [27] established a sensitive predictive algorithm to correctly distinguish between sporadic MSS and MSI-H CRCs using a 14 miRNA signature in 39 samples. Schepeler et al. [28] used microarrays to profile the expression of 315 miRNAs in normal mucosa samples (n=10) and stage II colon cancers (n=49) differing with regard to microsatellite status.
and recurrence of disease. They observed that miR-145 expression was lower in cancer relative to normal tissue and that microsatellite status could be correctly predicted based on miRNA expression profiles. In conclusion, although largely limited to sporadic MMR-deficient CRCs, current data provide evidence that perturbed expression of miRNAs in colon cancer may not only have a functional effect on tumor cell behaviour, but that some miRNAs with prognostic potential could be of clinical importance [29].

The results of this study will enhance the scarce knowledge available to date on the role of miRNAs in Lynch syndrome-associated colorectal cancer development. The findings are likely to shed light on intestinal carcinogenesis in general and allow further characterisation of the various pathways involved in MSI-related tumorigenesis. The identification and characterisation of miRNA expression signatures in Lynch syndrome cancer patients may not only improve clinical risk assessment and prognosis, but also offer an opportunity to identify novel therapeutic strategies for cancer patients.
References


Table 1: Top 20 miRNAs differentially expressed in MMR deficient compared to MMR proficient colorectal cancers cell lines

<table>
<thead>
<tr>
<th>miRNA Name</th>
<th>Direction</th>
<th>Pool1 count (MMR-)</th>
<th>Pool1 frequency</th>
<th>Pool2 count (MMR+)</th>
<th>Pool2 frequency</th>
<th>Log(Psame/Pdiff)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-499-5p</td>
<td>↓</td>
<td>1642</td>
<td>0.0010</td>
<td>119032</td>
<td>0.0090</td>
<td>-150763</td>
</tr>
<tr>
<td>hsa-miR-371-5p</td>
<td>↓</td>
<td>18184</td>
<td>0.0010</td>
<td>109922</td>
<td>0.0080</td>
<td>-99962</td>
</tr>
<tr>
<td>hsa-miR-200a</td>
<td>↓</td>
<td>350694</td>
<td>0.01</td>
<td>342852</td>
<td>0.025</td>
<td>-84781.8</td>
</tr>
<tr>
<td>hsa-miR-373</td>
<td>↓</td>
<td>14098</td>
<td>0.0010</td>
<td>81549</td>
<td>0.0060</td>
<td>-73114.1</td>
</tr>
<tr>
<td>hsa-miR-200b</td>
<td>↓</td>
<td>704912</td>
<td>0.019</td>
<td>479924</td>
<td>0.035</td>
<td>-56962.2</td>
</tr>
<tr>
<td>hsa-miR-372</td>
<td>↓</td>
<td>10283</td>
<td>0.0010</td>
<td>59112</td>
<td>0.0050</td>
<td>-52861.9</td>
</tr>
<tr>
<td>hsa-miR-224</td>
<td>↓</td>
<td>41743</td>
<td>0.0020</td>
<td>85516</td>
<td>0.0070</td>
<td>-46316.7</td>
</tr>
<tr>
<td>hsa-miR-371-3p</td>
<td>↓</td>
<td>6588</td>
<td>0.0010</td>
<td>44572</td>
<td>0.0040</td>
<td>-41790.8</td>
</tr>
<tr>
<td>hsa-miR-429</td>
<td>↓</td>
<td>152804</td>
<td>0.0040</td>
<td>156215</td>
<td>0.012</td>
<td>-41021.6</td>
</tr>
<tr>
<td>hsa-miR-141</td>
<td>↑</td>
<td>1762050</td>
<td>0.046</td>
<td>401587</td>
<td>0.029</td>
<td>-38202.7</td>
</tr>
<tr>
<td>hsa-miR-10a</td>
<td>↑</td>
<td>1986950</td>
<td>0.052</td>
<td>467442</td>
<td>0.034</td>
<td>-38149.1</td>
</tr>
<tr>
<td>hsa-miR-31</td>
<td>↓</td>
<td>602096</td>
<td>0.016</td>
<td>370715</td>
<td>0.027</td>
<td>-32469.1</td>
</tr>
<tr>
<td>hsa-miR-100</td>
<td>↑</td>
<td>164399</td>
<td>0.0050</td>
<td>6491</td>
<td>0.0010</td>
<td>-31483.3</td>
</tr>
<tr>
<td>hsa-miR-125a-5p</td>
<td>↓</td>
<td>117214</td>
<td>0.0040</td>
<td>119074</td>
<td>0.0090</td>
<td>-30921.8</td>
</tr>
<tr>
<td>hsa-miR-24</td>
<td>↓</td>
<td>566033</td>
<td>0.015</td>
<td>347730</td>
<td>0.026</td>
<td>-30194.2</td>
</tr>
<tr>
<td>hsa-miR-424</td>
<td>↓</td>
<td>30682</td>
<td>0.0010</td>
<td>56459</td>
<td>0.0050</td>
<td>-28132.9</td>
</tr>
<tr>
<td>hsa-miR-148a</td>
<td>↑</td>
<td>303974</td>
<td>0.0080</td>
<td>35600</td>
<td>0.0030</td>
<td>-26735</td>
</tr>
<tr>
<td>hsa-miR-125b</td>
<td>↑</td>
<td>125963</td>
<td>0.0040</td>
<td>5836</td>
<td>0.0010</td>
<td>-22502</td>
</tr>
<tr>
<td>hsa-miR-499-3p</td>
<td>↓</td>
<td>241</td>
<td>0.0010</td>
<td>16837</td>
<td>0.0020</td>
<td>-21238</td>
</tr>
<tr>
<td>hsa-miR-192</td>
<td>↓</td>
<td>627758</td>
<td>0.017</td>
<td>348276</td>
<td>0.026</td>
<td>-20761.9</td>
</tr>
</tbody>
</table>
8.2 Identification of novel recurrent duplication “hot spots” in Lynch syndrome colorectal cancers.

My contribution to this work:
- DNA extraction from Lynch syndrome and sporadic colorectal cancers;
- Analysis of high resolution chip array;
- Selection of candidate genes;
- Validation of selected candidate by qPCR;
- Statistical comparison;
- Manuscript preparation;
Identification of novel recurrent duplication “hot spots” in Lynch syndrome colorectal cancers

Salvatore Piscuoglio¹, Nicole Lüscher¹, Anamaria Dumea¹, Michal Kovac², Luigi Terracciano² and Karl Heinimann¹$.

¹Research Group Human Genetics, Department of Biomedicine, University of Basel, and Division of Medical Genetics, University Children’s Hospital, Basel, Switzerland.

²Institute of Pathology, University Hospital of Basel, Basel, Switzerland.
Abstract

Background

Lynch syndrome (also Hereditary Non-Polyposis Colon Cancer, HNPCC) represents the most common, autosomal dominantly inherited cancer predisposition worldwide and accounts for 3-5% of the total colorectal cancer (CRC) burden. It is caused by germline mutations in DNA mismatch repair (MMR) genes (mainly MLH1 and MSH2). MMR deficiency results in microsatellite instability (MSI), i.e. genome-wide accumulation of somatic alterations at repetitive DNA sequence motifs, and present in 90% of Lynch syndrome-related cancer. About 80% of MSI tumors have a near-diploid karyotype which stands in clear contrast to the microsatellite stable (MSS) cancers which predominantly are aneuploidy. Thus far, little is known on the type and frequency of, microdeletions and microduplications in LS-related CRCs.

Methods

Here, we applied a high-density CGH microarray-based method using the Affymetrix Whole Genome 2.7 M chip, to study somatic copy number aberrations in CRCs from 12 unrelated Swiss LS patients, and whose cancers displayed microsatellite instability with confirmed germline mutation in MLH1 or MSH2. Next we validated the results in 46 LS-related as well as 50 colorectal cancers by quantitative real-time PCR using locus specific primer pairs.

Results

Copy number assessment by CGH array revealed 2 novel somatic microduplications "hot spot" regions containing ERCC2 (19q13.32) and STK40 (1p34.3) gene. The frequency of these copy number aberrations were further validate in our cohort of Lynch syndrome cancers. CNA were present in 24/46 (60.8%) and in 18/46 (39.1%) CRCs for ERCC2 and STK40, respectively.

Conclusion

Our study identified novel microduplications in Lynch syndrome related CRC located on 1p34.3 and 19q13.32 These genes associated with CN changes in LS-related CRCs warrant further investigation to establish their possible clinical implications. Currently, to confirm that these CNA are an only present in Lynch syndrome CRC 50 sporadic CRCs are under investigation.
Introduction

Hereditary nonpolyposis colorectal cancer (HNPCC), also referred to as the Lynch syndrome (LS), with an estimation between 1:200 and 1:2000 predisposition is the most common form of autosomal dominantly inherited cancer predisposition worldwide. It is characterized by the occurrence of early onset colorectal carcinoma (CRC) as well as a distinct spectrum of extracolonic tumors, such as endometrium, stomach, ovarian, breast and renal pelvis cancers [1, 2] and caused by a germline mutation in mismatch repair (MMR) genes, MLH1 and MSH2 (90%) and MSH6 (10%) [1, 2]. Tumors with MMR deficiency exhibited frequent errors in microsatellite DNA, short segments of DNA containing tandem repeats of mono-, di-, tri- or tetranucleotide [3]. The high-frequency MSI (MSI-H) CRCs have unique clinicopathologic features, such as right-sided, mucinous or poorly differentiated, and stable chromosomal status in the tumors [4].

DNA copy number variation (CNV) and other structural variations in the human genome are increasingly recognized as an alternative source of genetic variation that may influence cancer risk and are a common occurrence in all forms of cancer [5-9]. A typical cancer sample exhibits an average of 17% amplifications and 16% deletions within an entire genome [10]. Somatic copy number alterations have been shown to significantly affect pathways involving tumor suppressor genes such as TP53, APC, BRCA1, BRCA2, PTEN, and RB1 and oncogenes, including HRAS and RET [5, 11]. Detection of these alterations and identification of the specific genes responsible for cancer proliferation can help to subtype cancers at the molecular levels and lead toward more individualized cancer-type specific therapies [12-15].

About 80% of MSI tumors have a near-diploid karyotype and a genetic alterations different from those of microsatellite stable (MSS) cancers [16-20]. Despite the progress of our understanding of CRC genetics, genomic alterations of various subtypes of CRC have not been fully characterized. Copy number variations (CNVs) can contribute to variable levels of gene expression [21], and thus fine-scale copy number (CN) profiling of cancer may further enhance our knowledge about tumorigenesis. This study aim to investigate 12 Lynch syndrome-related CRCs with high-density CGH microarray-based method in search of somatic microduplications and/or microdeletions.
Materials and methods

Patient characteristics
For this study a cohort of 52 Lynch syndrome-related colorectal cancers with identified germline mutation (30 MLH1, 16 MSH2, 3 MSH6, 2 PMS2) were analysed for CNV.

In addition, based on the recommendations of the National Cancer Institute workshop on MSI, a panel of microsatellite loci (BAT25, BAT26, D2S123, D5S346, D17S250) [3] and two additional microsatellite markers (BAT40, MYCL1) were used to determine MSI status.

Patient data including full follow-up were obtained by retrospective analysis of medical records, regional tumor registries and/or treating physicians. Tissue samples were obtained by surgical or endoscopic excision.

Genomic DNA Isolation
To isolate genomic DNA from fresh frozen tumor tissue samples the QIAamp DNA Mini Kit (Qiagen, Hombrechtikon, Switzerland) and for formalin fixed paraffin embedded (FFPE) tumor samples the RecoverAll™ Total Nucleic Acid Isolation Kit (Ambion, Invitrogen, Carlsbad, CA, USA) were used according to the manufacturers' guidelines.

Cytogenetics Whole-Genome 2.7M Array
A total of 500 ng of genomic DNA of six LS patients (each of them were matched with its normal counterpart), were analyzed using an Affimetrix Cytogenetics Whole-Genome 2.7M Array, a high resolution array containing approximately 400,000 SNP markers and 2.3 million non-polymorphic markers, with high density coverage across cytogenetically significant regions according to the manufacturer's instructions. Data was collected using either GeneChip® Scanner 3000 Dx and CEL files were analyzed using Affymetrix Chromosome Analysis Suite software (ChAS v.1.1). The annotation file used in our analysis can be found on the Affymetrix website, listed as ArrayNA30.2 (hg18). CNVs detected were compared with the Database of Genomic Variants (http://projects.tcag.ca/variation) for overlap with known copy number variants using previously described criteria [24].
Determination of selected copy number aberrations by Real-time PCR

Oligonucleotide primers for quantitative PCR (qPCR) were designed for each gene using the AlleleID software (PREMIER Biosoft International, Palo Alto, CA) (ERCC2 for: 5’-TGGGAAATGAAACGGGAACAG-3’; rev: GGGCAAGACGGACTACGG; STK40 for: 5’-GTCTCTGTTCCTGTCTC-3’; rev: GGCTGCGTAATATGATGG-3’; CFTR for: 5’- GCATGGGAGGAATAGGTGAA-3’; rev: 5’-CACAATCTACACAATAGGACATGG-3’), to assure maximal efficiency and sensitivity according to the following parameters: avoidance of the formation of self and hetero-dimers, hairpins and self-complementarity, primer length and melting temperature. These properties were further verified using different internet-based interfaces such as Primer-3 [25]. Melting curve analysis was always performed at the end of each PCR assay to control for specificity.

qPCR was performed using standard protocols with 2X iQ SYBR Green supermix (Bio-Rad) on a Bio-Rad iCycler. Briefly, 50 ng DNA was added to 12.5 µl of SYBR-green PCR master mix (Bio-Rad), with 0.5 µl (600 nM) of each primer, and water to a final volume of 25 µl. The reactions were amplified in a single step of 3 min at 95°C and then for 40 cycles of 10 s at 95°C, 1 min at 60°C, with final denaturation step for 10 sec at 95°C. The thermal denaturation protocol was run at the end of the PCR to determine the number of products that were present in the reactions. Experiments were done in triplicate and included non-template controls for each gene. The amount of each gene was normalized to cystic fibrosis transmembrane conductance regulator (CFTR) as reference gene. We used a standard analysis to calculate the amplification of the genes by the $2^{\Delta\Delta Ct}$ method as described previously [26]. Results for each sample were expressed as the N-fold copy number change [27].
Results
Initially we investigated six LS-related CRCs from 4 known MLH1 and 2 MSH2 germline mutation carriers matched with their tumor free counterpart for the presence of copy number aberrations (CAN), i.e. micro- deletion/duplication using the Affymetrix Whole-Genome 2.7M CGH array.
We observed somatic gains (microduplications) at 1p34.3, 2q24.3, 3p23, 5q23.2, 5q33.1, 6p23, 8q11.1, 10p15.1, 10q25.1, 12p11.23, 17p13.1, 17q24.3, 19q13.32, Xp22.2 and somatic losses (microdeletions) at 19p12, 3p14.2. The overall findings of DNA copy number gains and losses across all samples are shown in Table 1. Interestingly, we identified 2 recurrent novel somatic microduplicated “hot spots” for genomic rearrangements located on 1p34.3 and 19q13.32 containing the ERCC2 and STK40 gene, respectively (Figure 1 A and 1 B).
To further assess the frequency at these loci we evaluated 46 LS-related CRCs from 30 MLH1, 16 MSH2, 3 MSH6 and 3 PMS2 mutation carriers by qPCR. The analysis revealed microduplications in ERCC2 were present in 24/46 (60.8%) CRCs and in STK40 in 18/46 (39.1%) CRCs (Table 2; Figure 2 A and 2 B).
In order to assess if these novel recurrent “hot spot” region represent somatic alterations typical for LS-related tumorigenesis, a set of 50 sporadic, microsatellite-stable CRCs is currently under investigation.
Brief discussion
The understanding of chromosomal aneuploidies and their role in tumor development is a fundamental problem in cancer biology. Chromosomal aneuploidy, the gain or loss of chromosomes is the most common alteration in cancer. The majority of cancer cells in sporadic MSS CRCs have numerical and structural chromosomal abnormalities with translocations, deletions and other aberrations. In contrast, about 80% of MSI tumors display a near-diploid karyotype and a distinct genetic alteration distinguishable from those observed in MSS cancers [19]. In this study, using the Affymetrix Whole Genome 2.7 M chip array in 6 fresh-frozen Lynch syndrome-associated CRCs we identified 2 novel, recurrent microduplication “hot spot” regions on chromosomes 1p34.3 and 19q13.32 containing the ERCC2 and STK40 genes, respectively, and validated the findings by qPCR in a cohort of 46 additional LS-related CRCs. To assess if these alterations are specific to LS related cancers currently 50 sporadic CRCs is currently under investigation.

ERCC2 is a key component of the nucleotide excision repair pathway. Further, the protein is an integral member of the basal transcription factor BTF2/TFIH complex, displays an ATP-dependent DNA helicase activity and belongs to the RAD3/XPD subfamily of helicases [28]. Defects in this gene can result in three different disorders: the cancer-prone syndrome xeroderma pigmentosum complementation group D, trichothiodystrophy, and Cockayne syndrome. Little is known concerning the STK40 that may be a negative regulator of NF-kappa-B and p53-mediated gene transcription.

To date, it remains to be clarified if and how these microduplications and amplifications (up to 6n) may affect cancer cell proliferation and progression in Lynch syndrome colorectal cancers.

Among all somatic mutations, non-germline CNVs found in the cancer genomes, also known as copy number alterations/aberrations (CNAs), are frequently observed, e.g., gains of oncogene and losses of tumor suppressor gene loci [22]. Furthermore, the DNA CN states of CRC cases are related to the response of drug treatments, e.g., the degree of CRC-related CNA is associated with response to systemic combination chemotherapy with capecitabine and irinotecan [23].
Figures and Tables

Figure 1: ChAS analysis of LS-related CRCs

Figure 1: A) Minimal region duplicated on chr19 (19q13.32) including ERCC2 gene in 3 different patients. B) Minimal region duplicated on chr1 (1p34.3) including STK40 gene in 2 different patients.
Figure 2: qPCR results in 46 LS-related CRCs: A) CNAs in ERCC2 gene. B) CNAs in STK40 gene.
<table>
<thead>
<tr>
<th>Gene name</th>
<th>Copy number aberrations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2n (diploid)</td>
</tr>
<tr>
<td>ERCC2</td>
<td>22</td>
</tr>
<tr>
<td>STK40</td>
<td>28</td>
</tr>
</tbody>
</table>

**Table 2**: Degree of copy number aberrations in 46 LS-related CRCs
<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Locus</th>
<th>Alteration</th>
<th>Base Start</th>
<th>Base End</th>
<th>Del Size (Mb)</th>
<th>Patient n.</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1p34.3</td>
<td>Duplication</td>
<td>36583571</td>
<td>36636982</td>
<td>53.4</td>
<td>5</td>
<td><strong>STK40, LSM10</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>36567450</td>
<td>36623262</td>
<td>55.8</td>
<td>7</td>
<td><strong>STK40</strong></td>
</tr>
<tr>
<td>2</td>
<td>2q24.3</td>
<td>Duplication</td>
<td>168993733</td>
<td>169032679</td>
<td>38.9</td>
<td>1</td>
<td><strong>LASS6</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>168993733</td>
<td>169036109</td>
<td>42.3</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3p23</td>
<td>Duplication</td>
<td>31535493</td>
<td>31557397</td>
<td>21.9</td>
<td>7</td>
<td><strong>STT3B</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>31540540</td>
<td>31557397</td>
<td>16.8</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3p14.2</td>
<td>Deletion</td>
<td>60421679</td>
<td>60439908</td>
<td>18.2</td>
<td>11</td>
<td><strong>FHIT</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>60430695</td>
<td>60450766</td>
<td>20</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>60530263</td>
<td>60542710</td>
<td>12.4</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>60497433</td>
<td>60554225</td>
<td>56.7</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5q23.2</td>
<td>Duplication</td>
<td>122869729</td>
<td>122896827</td>
<td>27</td>
<td>11</td>
<td><strong>CSNK1G3</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>122871101</td>
<td>122896827</td>
<td>25.7</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5q33.1</td>
<td>Duplication</td>
<td>151109899</td>
<td>151145327</td>
<td>35.4</td>
<td>5</td>
<td><strong>G3BP1</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>151131753</td>
<td>151142191</td>
<td>10.4</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6p23</td>
<td>Duplication</td>
<td>15406555</td>
<td>15428655</td>
<td>22.1</td>
<td>5</td>
<td><strong>JARID2</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15402243</td>
<td>15454282</td>
<td>52</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8q11.1</td>
<td>Duplication</td>
<td>47134349</td>
<td>47261260</td>
<td>126.9</td>
<td>1</td>
<td>Intragenic region</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47120825</td>
<td>47290072</td>
<td>169.2</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10p15.1</td>
<td>Duplication</td>
<td>6168193</td>
<td>6192155</td>
<td>23.9</td>
<td>11</td>
<td><strong>RBM17</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6168193</td>
<td>6192694</td>
<td>24.5</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10q25.1</td>
<td>Duplication</td>
<td>111746435</td>
<td>111756922</td>
<td>10.4</td>
<td>1</td>
<td><strong>ADD3</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>111747464</td>
<td>111769074</td>
<td>21.6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>12p11.23</td>
<td>Duplication</td>
<td>26952817</td>
<td>26993684</td>
<td>40.8</td>
<td>16</td>
<td><strong>FGFR1OP2</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26974916</td>
<td>26993684</td>
<td>18.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>17p13.1</td>
<td>Duplication</td>
<td>10499352</td>
<td>10536740</td>
<td>37.3</td>
<td>7</td>
<td><strong>SCO1, MYH3</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10528556</td>
<td>10566728</td>
<td>38.1</td>
<td>11</td>
<td><strong>SCO1</strong></td>
</tr>
<tr>
<td>17</td>
<td>17q24.3</td>
<td>Duplication</td>
<td>64815529</td>
<td>64839814</td>
<td>24.2</td>
<td>1</td>
<td><strong>MSI2, ABCA5</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>64827465</td>
<td>64848873</td>
<td>21.4</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>19q13.32</td>
<td>Duplication</td>
<td>50517505</td>
<td>50549160</td>
<td>31.6</td>
<td>1</td>
<td><strong>ERCC2, KLC3, CKM</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50539800</td>
<td>50550532</td>
<td>10.7</td>
<td>5</td>
<td><strong>ERCC2, KLC3</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50540437</td>
<td>50669798</td>
<td>129.3</td>
<td>5</td>
<td><strong>ERCC2, KLC3, PPP1R13L, CD3EAP, ERCC1</strong></td>
</tr>
<tr>
<td>19</td>
<td>19p12</td>
<td>Deletion</td>
<td>22987786</td>
<td>23041491</td>
<td>53.7</td>
<td>11</td>
<td><strong>MUC16</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>22987786</td>
<td>23041491</td>
<td>53.7</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>Xp22.2</td>
<td>Duplication</td>
<td>9585012</td>
<td>9610052</td>
<td>25</td>
<td>3</td>
<td><strong>TBL1x</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9591311</td>
<td>9623919</td>
<td>32.6</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1:** CGH analysis results for copy number aberration in 6 LS-related CRCs.
References


9. REFERENCES

1. (IARC), G., Section of Cancer Information 2010.


61. Boland, C.R., et al., *A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of


188. Bandic, D., et al., *Expression and possible prognostic role of MAGE-A4, NY-ESO-1, and HER-2 antigens in women with relapsing invasive ductal breast


10. CURRICULUM VITAE

PERSONAL INFORMATION

Name                          Salvatore Piscuoglio
Address                       Strassburgeralle , 29, 4055 Basel, Switzelan
Telephone                     +41774303611
E-mail                        s.piscuoglio@unibas.ch - s.piscuoglio@gmail.com

Nationality                   Italian

Date of birth                 23/02/1982

WORK EXPERIENCE

• Dates (from – to)           01-01-2009 – till now
• Name and address of        Department of Biomedicine (University of Basel),
  employer                     Mattenstrasse 28, CH-4058 Basel.
• Type of business or        Translational research
  sector
• Occupation or position     PhD Student (Human Genetics group, PD Karl
  held                        Heinimann)
• Main activities and         MicroRNA, DNA, RNA extraction and sequencing,
  responsibilities            library generation and next generation sequencing,
                                microsatellite analysis, Sanger sequencing, statistical
                                analysis, immunochemistry, tissue microarray
                                evaluation, long range PCR, real time PCR, MLPA.
WORK EXPERIENCE

• Dates (from – to) 01-04-2008 – 31-12-2008
• Name and address of employer Institute of Pathology (University of Basel), Division of Molecular Pathology (chief Luigi M.Terracciano), Schönbeinstrasse 40. CH-4031
• Type of business or sector Clinical research
• Occupation or position held Research Fellow
• Main activities and responsibilities Tissue micro array building, FISH, RNA and DNA extraction, ISH, microscopy analysis, statistical analysis.

WORK EXPERIENCE

• Dates (from – to) 04/11/2002 - 21/2/2008
• Name and address of employer Department of Biomorphological and Functional Sciences, Section of Pathology, University of Naples "Federico II", Naples, Italy.
Department of Molecular and Cellular Pathology “L.Califano” University of Naples "Federico II”).
• Type of business or sector Clinical research
• Occupation or position held Research Fellow
• Main activities and responsibilities Static analysis, DNAploidy, immunohistochemistry reaction, special stain reactions, macroscopic and microscopic analysis, immunofluorescence, PCR, Whole Mount reaction, In situ Hybridization, Rt-Pcr, Fish, Western Blotting, Southern Blot, Mini and Maxi prep.
EDUCATION AND TRAINING

• Dates (from – to) 20/12/2005 - 26/03/2008
• Name and type of organisation providing education and training University of Naples “Federico II” Faculty of Biotechnological Science Medical (curriculum)
• Principal subjects/occupational skills covered Molecular biology, genetics, gene therapy, cellular biology
• Title of qualification awarded Master of Science

Thesis (MSc): “Chromatin Assembly Factor-1, a marker of prostate cancer progression?”

• Dates (from – to) 19/09/2001 - 20/12/2005
• Name and type of organisation providing education and training University of Naples “Federico II” Faculty of Biotechnological Science
• Principal subjects/occupational skills covered Biochemistry, molecular and genetic basis of diseases, pathology.
• Title of qualification awarded Bachelor of Science

Thesis (BSc): “Morphometric evaluation of chromatin texture and CAF-1 expression in oral cancer”
PERSONAL SKILLS 
AND COMPETENCES

MOTHER TONGUE        ITALIAN

OTHER LANGUAGES      English

• Reading skills       Fluent
• Writing skills       Fluent
• Verbal skills        Fluent

SOCIAL SKILLS 
AND COMPETENCES
I’ve worked several times in projects where collaboration with other members of the group was essential; furthermore, for 6 years (2000-2006) I got experience as tourist animator and club manager.

ORGANISATIONAL SKILLS 
AND COMPETENCES
When I was the superintendent of a staff of 15 people I managed to arrange their jobs and activities, recently I’ve worked as a responsible for human resources of a magazine.

TECHNICAL SKILLS 
AND COMPETENCES
I have good experience in using the main laboratory’s tools and I have an excellent familiarity in using computers (Windows, Mac, Ms Office, Macromedia Studio, SPSS for statistic analysis, Photoshop, Statview.)

ARTISTIC SKILLS 
AND COMPETENCES
I’m Sound engineer and Dj, and I play guitar
List of Publications:


Manuscript submitted or in preparation:


different classes of copy number mutation in colorectal cancer gene. Manuscript in preparation

7. Luca Quagliata*, Mariacarla Andreozzi*, Michal Kovac, Luigi Tornillo, Zuzanna Makowska, Marcus Heim, Karl Heinimann, Salvatore Piscuoglio$ and Luigi Terracciano$. **SH2D4A is frequently downregulated in hepatocellular carcinoma.** Manuscript in preparation (equally contributed, corresponding author)


11. ACKNOWLEDGMENTS

It is a pleasure to thank the many people who made this thesis possible.

It is difficult to overstate my gratitude to my Ph.D. supervisor, Dr. Karl Heinimann. With his enthusiasm, his inspiration, and his great efforts to explain things clearly and simply, he helped to make research fun for me. Throughout my PhD period, he provided encouragement, sound advice, good teaching, good company, and lots of good ideas. It was also through his persistence, understanding and kindness that I completed this important step of my life. I would have been lost without him.

I would like to express my profound gratitude and thank to Prof. Luigi Terracciano that during this time has been like a father. He is the one who truly made a difference in my life. He provided me with direction, technical support and became more of a mentor and friend, than a professor. I doubt that I will ever be able to convey my appreciation fully, but I owe him my eternal gratitude.

I would also like to warmly thank my laboratory friends Marianne, Michele, Sybille and Nemya, for their technical advice and assistance, their never ending patience and of course, for the fun times we had together. Every research team should have a people like them!!!

I’d like to thanks my PhD father prof. Primo Schär for his thesis supervision and for his advises during this journey; his wisdom, knowledge and commitment to the high standards inspired and motivate me.

I would like to express my heartfelt gratitude to Professor Alfredo Fusco and Dr. Pierlorenzo Pallante for their expertise on HMGA proteins and their sincere friendship.
I would also thank Dr. Shivendra Kinshore and Prof. Mihaela Zavolan for their hard collaborations, their expertise in computational and RNA biology made possible the development of many future projects.

Thanks to Dr. Michal Kovac, that during these years he helped me to solve critical problems, writing, giving wise advice, and helping with various applications.

I am indebted to my many colleagues/friends for providing a stimulating and fun environment in which to learn and grow. I am especially grateful to Maricarla, Anamaria, Francesca, Luigi, Luca, Giulio, Markus and Ezia.

I wish to thank my friends Giulia, Mike, Elena, Jenna, Ming, Claus, Daniela, Giancarlo, Meghan, Francesca A. and Luigi for helping me get through the difficult times, and for all the emotional support, entertainment, and caring they provided.

Lastly, and most importantly, I wish to thank my sister Giusy and my parents Carmela and Gaetano. They bore me, raised me, supported me, taught me, and loved me. To them I dedicate this thesis.